DNA Replication Termination and Chromosome Dynamics in *Escherichia coli*

A thesis submitted

for the degree of Doctor of Philosophy

***by* Daniel J. Goodall**

Email: Daniel.Goodall@brunel.ac.uk

Department of Biosciences,

College of Health, Medicine and Life Sciences,

Brunel University London,

London,

Uxbridge,

UB8 3PH

Contents

Acknowledgments

I must thank my mentor, Christian J. Rudolph, for his invaluable guidance throughout this project. Our engaging discussions about biology have been incredibly enriching. I will certainly miss those chats. I also thank Juachi Dimude for early lab instruction and later troubleshooting sessions as a colleague.

I would like to express my profound gratitude to my colleagues for their priceless camaraderie and dynamic exchanges. You guys really made this doctoral journey less of a solitary trek and more of a team adventure. Cheers to you all!

Massive thanks to my family - my ever supportive parents and grandparents for always encouraging me forward while being proud no matter what.

I dedicate this to my late grandparents, Chris and Ted, who were instrumental in motivating me to persist and I would not be on this path without them. I thought of them constantly.

To my partner, who was there in the toughest late stages, giving me energy to continue, reminding me to focus on what I can control and letting me talk through problems to find solutions when I came up against tough hurdles - you're the best.

Thanks to my friends for needed distraction over a beer every now and then, and to my Jiu Jitsu team at Jay Butler BJJ for a home on the mats throughout this whole thing, it helped keep my mind clear of frustrations and stay balanced.

Finally, I am truly grateful for this research opportunity that has taught me too many life lessons to list, and what it means to be a true scientist, following the data wherever it leads.

Declaration

The research presented in this thesis is my own work, unless otherwise specified, and has not been submitted for any other degree. - *Daniel J. Goodall*

Abstract

Faithful genome duplication requires accurate replication initiation, elongation, and termination. In *Escherichia coli*, replication initiates bidirectionally from *oriC* and concludes when converging forks meet in the termination region. Recent work from the Rudolph lab has shown that this fusion process is risky and can jeopardize stability of the termination area. *E. coli*' s chromosome contains a specialized "replication fork trap" comprised of Tus binding *ter* sites to direct termination events into a confined region. However, the function and evolution of this fork trap system is unclear, as are the key proteins and pathways needed to complete replication. This work investigates chromosome architecture and impacts of fork fusions on genomic instability in *E. coli* by measuring recombination at defined chromosome locations and accumulation of R-loops that threaten genomic stability. Bioinformatics revealed maintenance of the fork trap structure across diverse *E. coli* strains and related enterobacteria, and that inactivation of Tus leads to mild growth defects, showing that while not all bacteria have a fork trap system, it is maintained once acquired leading to advantageous growth. Ectopic replication fusion events increased local recombination rates, implying fork collisions threaten genomic stability, with RecG and UvrD mutants showing synergistic effects of DNA metabolism in the termination area. R-loop analysis demonstrated Tus synergizes with accessory helicases RecG and UvrD to limit global R-loop accumulation, and I suggest the R-loop toxicity begins in the termination area. Unrestrained R-loop buildup induced toxicity over time, highlighting the danger of unrestricted R-loop accumulation and importance of enzymatic processing by RecG and UvrD. These findings provide key insights into replication termination in *E. coli*. They suggest the fork trap and repair pathways helps contain deleterious consequences of fork fusions, while promoting proper chromosome segregation. This work establishes principles relevant to both prokaryotes and eukaryotes regarding maintenance of genomic integrity during the essential process of DNA replication. Elucidating bacterial systems that enable faithful completion of replication enhances fundamental understanding of chromosome duplication mechanisms required for genome stability across all domains of life.

Introduction

DNA replication is a process central to molecular biology which provides a foundation for all other cellular processes and pathways. From intracellular cell signalling pathways to cell-cell communication throughout an organism, the gene regulation of each organism dictates health and survivability (Kuzminov, 2016, 2018; Vassilev and DePamphilis, 2017). Prokaryotic cells proliferate through binary fission where a cell divides in two yielding two identical daughter cells. This process is called mitosis in eukaryotes. Therefore, for successful cell division to take place, the DNA content of the cell must be duplicated accurately to ensure as few alterations as possible in the daughter cell. This is said to maintain genome stability, a general term when referring to alterations in the genome that can lead to pathological consequences depending on which local region of the chromosomes are disrupted (Dimude et al., 2018a; Rudolph et al., 2013; Rudolph et al., 2019). In humans, dysregulation of gene expression can lead to conditions such as cystic fibrosis, Fredrich's ataxia and even cancer, showing the importance of researching the fundamental mechanisms of DNA replication and factors affecting gene regulation (Tomasetti et al., 2017; Venkitaraman, 2019; Willis et al., 2014). Studying DNA replication dynamics in eukaryotic systems is a complex and time-consuming task due to their intricate cellular structure and functions, including multiple chromosomes within a nucleus and various interacting organelles. Conversely, prokaryotic cells like *E. coli*, with their simpler structure and single circular chromosome, provide a more efficient alternative for such studies. Their rapid cell division allows for quicker observation of DNA replication and chromosome dynamics, offering valuable insights into the molecular mechanisms of replication. Therefore, while eukaryotic systems are important to study, using prokaryotes as a model system can provide a more time-effective approach to understanding DNA replication dynamics, paving the way for further research into more complex eukaryotic systems.

Escherichia coli (*E. coli*) has long been the organism of choice for research in molecular biology. Its rapid growth rate, ease of genetic manipulation, and the extensive knowledge of its biol-

ogy make it an ideal model organism for studying various biological processes, including DNA replication. The *E. coli* genome is a single circular chromosome, which simplifies the study of DNA replication. The replication process in *E. coli* starts at a single origin of replication (*oriC*) and proceeds bidirectionally until it reaches the terminus region (*ter*). This well-defined replication process allows for precise experimental manipulation and observation (Blattner et al., 1997). This is a Gram negative bacterium part of the wider Enterobacteriaceae family, also consisting of *Shigella*, *Klebsiella*, *Salmonella* and *Pseudoalteromonas*, to name a few. Bacteria are single cell organisms which package their genome into circular chromosomes and circular plasmids, which are smaller DNA elements often containing antibiotic resistance genes and other elements which can be shared across the population, a process termed horizontal gene transfer (HGT) (Pupo et al., 2000; Rasko et al., 2008; Touchon et al., 2009). Both the chromosome and plasmids can also contain toxin producing genes allowing for evolution of pathogenic strains to occupy particular niches. *E. coli* occupies the small intestine of most warm blooded animals including humans, cows, pigs and birds, but although some strains such as the well studied enterohaemorrhagic *E. coli* (EHEC) serotype O157:H7 pose serious health risks, others are commensal and do not cause disease. This highlights the variable niches that *E. coli* occupies. Understanding the molecular mechanisms at play in these bacteria can also offer new insights into novel antibiotic targets, such as those inhibiting the transcribing RNA polymerases or replicating DNA polymerases (Melton-Celsa, 2014; Zenkin et al., 2005).

The *E. coli* genome consists of a single circular chromosome of 4.6Mb and contains just over 4000 coding sequences. One distinguishing feature of bacteria is their high gene density and lack of pseudo genes. *E. coli* are a diverse species showing a core conserved genome of around 2000 genes, with a pangenome of around 18,000 genes (Sims and Kim, 2011; Touchon et al., 2009). Bacterial genomes vary greatly in size and structure depending on their ecological niches. They can be as small as 0.159 Mbp with 182 genes in *Carsonella ruddii*, an endosymbiont of psyllid insects (Nakabachi et al., 2006), or as large as 13 Mbp with 9376 genes in *Sorangium cellulosum*, a soil bacterium (Schneiker et al., 2007). Most bacteria have a single circular chromosome, but some have multiple chromosomes, such as_ Vibrio cholerae_ with two (Galli et al., 2019) and *Paracoccus denitrificans* with three (Winterstein and Ludwig, 1998). Some bacteria also have linear chromosomes or a combination of circular and linear chromosomes, such as *Agrobacterium tumefaciens*, showcasing the variety of chromosome arrangements in bacteria, despite the overly simplistic view of their genomes.

The *E. coli* chromosome is packaged into a 3D nucleoid, being associated with protein and

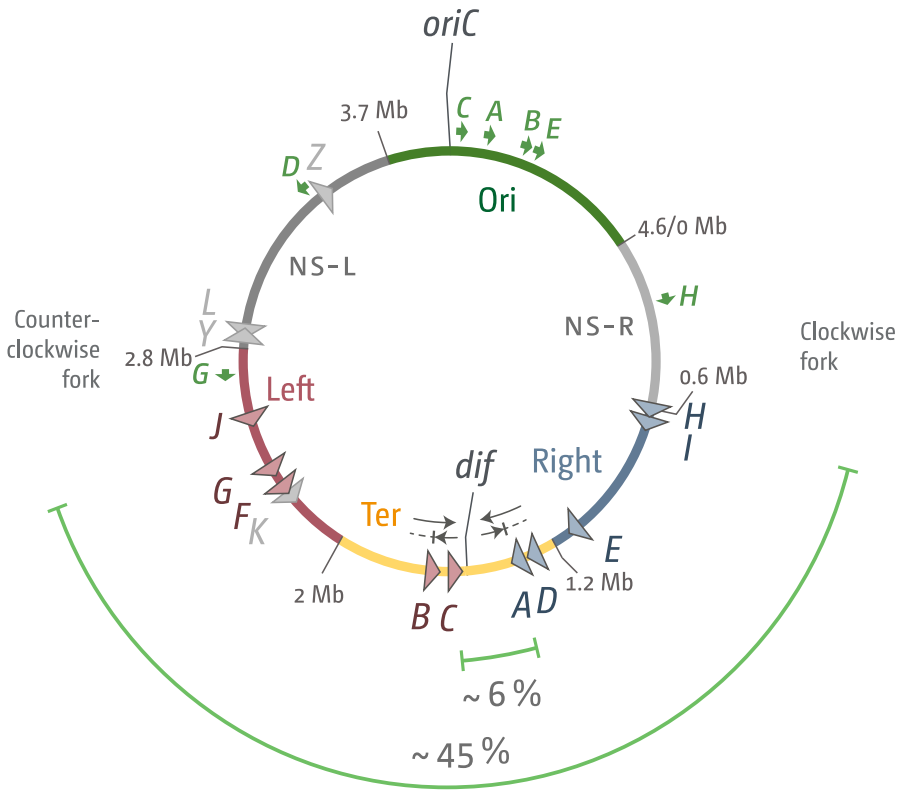

Figure 1: Chromosome map of *E. coli* showing macrodomains colour coded with green showing the Ori domain, red and blue showing Left and Right domains, respectively, and yellow showing the Ter domain. Grey non-structured (NS) domains are also shown. Polar oriented *ter* sites are shown as triangles pointing in the direction of permissibility, with red sites *terBCFGJ* arresting forks approaching from the right replichore and blue sites *terADEIH* arresting forks approaching from the left replichore. Greyed out *terKLYZ* show non-functional pseudo-sites. *rrn* operons are shown by green arrows (DG on the left and CABEH on the right replichores), *oriC* and *dif* resolution site are also shown. Green boundaries show the inner 'primary' frok trap consisting of *terABCD* at 6% the chromosome and the wider fork trap architecture covering 45% of the chromosome. Used from Goodall et al. 2021

RNA. The nucleoid is arranged by DNA supercoiling and protein interactions which separate the chromosome into macrodomains. The arrangement of these macrodomains are such that they align with metabolic processes, DNA replication and recombination (Reyes-Lamothe et al., 2008; Verma et al., 2019). In Figure 1 we can see that the chromosome can be grouped into functional macrodomains of Ori, Left, Right and Ter. The origin of replication in bacteria is identifiable through bioinformatics analysis of the genome utilizing the GC skew (Kono et al., 2011, 2012), and the Ter macrodomain was even defined through bioinformatics analysis of the *matS* site, which when bound by MatP helps to organise this area of the 3D nucleoid (Mercier et al., 2008). At the molecular level, nucleoid-associated proteins (NAPs) like HU (histone-like protein), compact and organize DNA by bending, looping, and restraining negative supercoils induced by DNA topology (Verma et al., 2019). Additionally, the macrodomain Ter protein MatP and structural maintenance protein MukB facilitate long-range DNA interactions within specific macrodomains. More recent research has also revealed that nucleoid DNA contacts the cell membrane to form an overall helical ellipsoid shape, which varies with growth phase due to changes in supercoiling, gene expression, and abundances of NAPs (Verma et al., 2019).

From Figure 2 we can see the spatiotemporal organisation of the *E. coli* chromosome during DNA replication and the final stages of cell division. In *E. coli*, genetic loci are spatially organized according to the genetic map, with the origin *oriC* localized near mid-cell and loci progressively further from *oriC* positioned farther from mid-cell. This spatial arrangement is a consequence of *oriC* anchoring near mid-cell after replication initiation, while replicated DNA is sequentially segregated to the cell quarters soon after synthesis. The temporal process of layered DNA segregation after replication establishes the organization of sister chromosomes that is then maintained through the rest of the cell cycle. As replication nears termination, the two replication forks move back towards mid-cell guided by *KOPS* sites so that FtsK, along with XerCD recombinase can mediate final unlinking and segregation of sister chromosomes can be completed prior to cell division. (Reyes-Lamothe et al., 2008)

The work by Sims and colleagues constructed two types of phylogenetic trees of *E. coli* and *Shigella* - a phenetic tree based on overall genomic composition reflecting phenotypic similarities of current organisms (A in Figure 3), and an evolutionary tree based on conserved core features representing ancestral relationships (B in Figure 3). They suggest that the ancestral origin of *E. coli* appears to be an opportunistic pathogen, based on the basal position of subgroup B2 in the evolutionary tree, suggesting pathogenicity is an ancestral trait that gave rise to commensal strains like subgroup A through a reductive evolutionary process. *Shigella* has emerged through

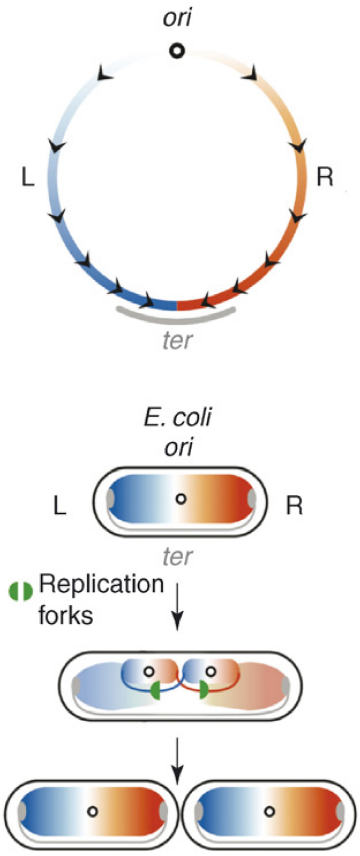

Figure 2: Spatiotemporal organization of the *E. coli* chromosome during DNA replication and cell division reveals sequential segregation of replicated DNA (Sourced from Reyes-Lamothe et al. (2008).

convergent evolution of two separate ancestor strains to share a pathogenic phenotype, underscoring the importance of lateral gene transfer in shaping phylogroup traits, as shown through the identification of distinguishing genomic features for each phylogroup (Sims and Kim, 2011). From the phylogenetic trees in Figure 3 we can see grouping of different *E. coli* strains which harbour commensal and pathogenic strains, we can also see *Shigella* and *Salmonella enterica* being close relatives. Some groups of *E. coli* are commensal and other cause enterohaemorrhagic disease (Picard et al., 1999; Sims and Kim, 2011), highlighting the range of niches this organism can occupy. More recent analysis shows even more divergence in the phylogeny of a particular virulent subset of *E. coli,* labelled now as group G (Clermont et al., 2019) and a more recent analysis using all available *E. coli* and *Shigella* genomes available on GenBank were compared and grouped into phylogroups using a Mash-based analysis, which extends our classification of this species of bacteria (Abram et al., 2021).

Considering DNA replication consists of 3 stages; initiation, elongation and termination, much of the scientific focus has been on the initiation or elongation stages of DNA replication, especially in eukaryotes. However, the final stage, termination, holds equal importance and complexity. The dynamics of termination, particularly the role of key proteins in regulating this stage, are critical to understanding the evolution of chromosome architectures and features. This is particularly evident when considering the consequences of fork fusions during termination, which are known to cause instability to the chromosome (Dimude et al., 2016; Rudolph et al., 2013; Rudolph et al., 2019). Research from the Rudolph lab has shown that when multiple origins are included on the chromosome, there is an increase in instability and gross chromosomal rearrangements occur as suppressors of these extra origins in order to maintain stability (Dimude et al., 2018a). This suggests a high selection pressure for more stable replisomes with high fidelity, as errors in replication can lead to a less stable chromosome. Therefore, a deeper understanding of the termination stage of replication, especially in organisms like *E. coli*, could provide valuable insights into chromosome stability and evolution.

In the context of DNA replication termination, *E. coli* provides a unique advantage, it has a specialized system for replication termination, involving a replication fork trap created by Tus-*ter* complexes. This system ensures that replication forks do not proceed past the terminus region, preventing head-on collisions between replication and transcription machinery. The simplicity and efficiency of this system make *E. coli* an excellent model for studying the intricacies of replication termination. However, it's important to note that while *E. coli* provides a simplified system for studying DNA replication, it may not fully represent the complexities of this process

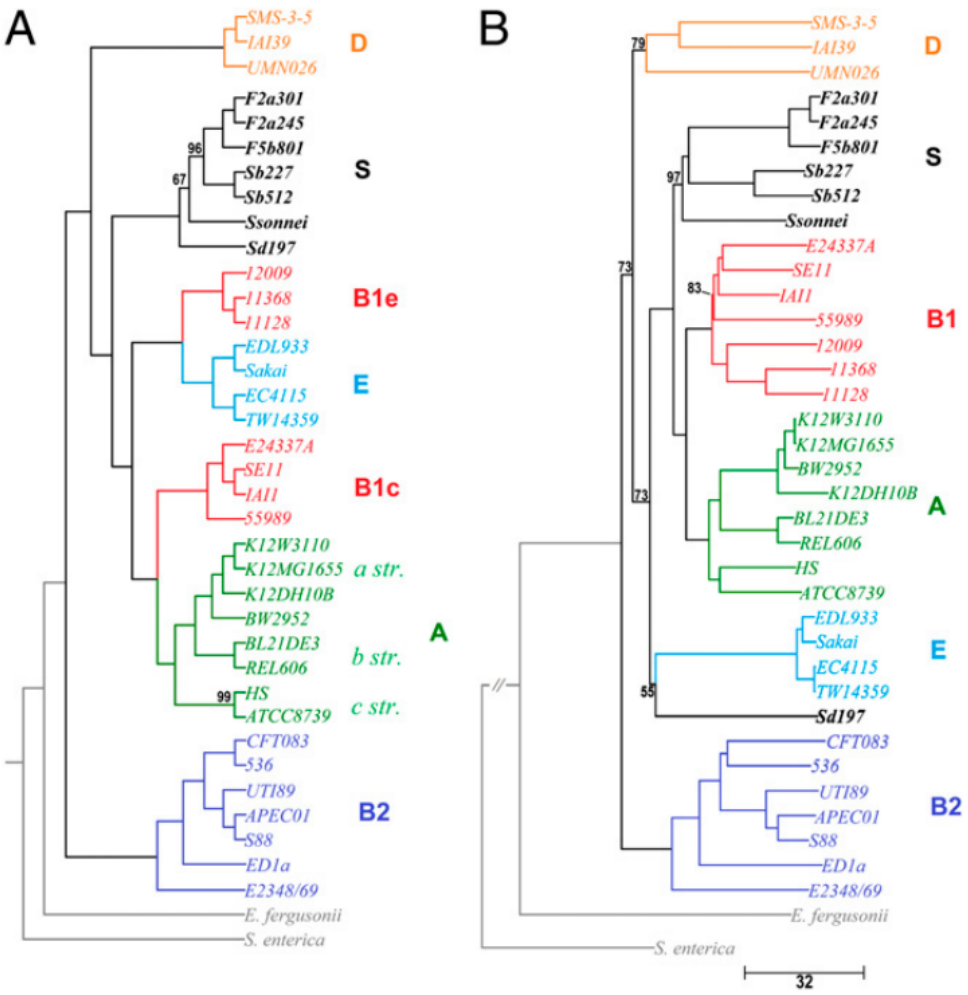

Figure 3: Phylogenetic tree of *E. coli* and *Shigella* strains (Repurposed from Sims 2011). Groups are colour coded and show strains belonging to commensal or pathogenic variants. A) Tree constructed from phenotypic similarities and B) Tree constructed from the ancestral lineage of each strain. Although differences can be observed, for instance B1 splitting into two further sub-groups when classified from phenotypic similarities, Sims et al. 2011 show a consistent grouping of the strains measured.

in eukaryotic organisms (Duggin and Bell, 2009; Rudolph et al., 2009). However, these simplified systems can be applied to eukaryotic systems. For example, the Willis group (Willis and Scully, 2016; Willis et al., 2014) has observed that the BRCA1 gene controls homologous recombination at replication forks held at Tus-*ter* arrays incorporated into the mammalian genome, showing that defects in this fork restart and repair increased sensitivity to cancer. This is a good example of how understanding bacterial replication systems can be applied to our understanding of cancer pathogenesis and other genetic conditions (Willis et al., 2017).

Over the years, multiple independent labs have deepened our understanding of DNA replication using *E. coli* as the model organism. Through genome manipulation, the Rudolph lab have shown that the addition of ectopic origins, *oriX* and *oriZ*, into the left and right hand replichores, respectively, pose significant challenges for the cell and disrupt the chromosome dynamics and symmetry of the duplicating chromosome (Dimude et al., 2018a; Massy et al., 1984). Changes in the replichore architecture by the addition of ectopic origins have been shown to induce gross chromosomal rearrangement (GCR) that act as suppressor mutations for the chromosomal changes. These changes often result in slower growth phenotypes and these GCR are thought to be a response to a change in the selection pressure to re-establish a similar replichore arrangement. The GCR are seen in two areas when ectopic *oriX* and *oriZ* are included but *oriC* is inactivated, leaving these new origins as the sole firing site that is still dependent on DnaA to replicate the chromosome. With *oriX* and *oriZ* there is a rearrangement of the terminus and *rrn* operons, respectively, with $oriX^{+}$ cells we see the rearrangement in the terminus area which inverts the right hand replichore *ter* sites, now in the permissible orientation, to allow forks coming from *oriX* to pass through into the right replichore and fuse with the opposing fork. In $oriZ^{+}$ cells there is an inversion of the *rrnCABEH* operons, which would minimize the amount of replication-transcription conflicts by re-establishing transcriptional co-directionality (Dimude et al., 2018a; Merrikh and Merrikh, 2018). These two GCR highlight the importance of *E. coli* chromosome architecture and illustrate that the chromosome is plastic and able to adapt to change when the selection pressure is great enough. The replication fork trap (RFT or fork trap), is a genetic system that plays a vital role in modulating replication fork dynamics in *E. coli*, and resides in the terminus region. At the heart of the RFT lies polar 23 bp *ter* sites, which serves as a binding platform for the monomeric protein Tus (termination utilisation substance). The Tus protein, through its polar binding to a *ter* site, establishes specific interactions with both the bases and sugar-phosphate backbone of the *ter* site, thereby influencing the permissibility of an approaching replication fork. The fact that the terminus area has been inverted in these strains highlights

the RFT as an important genomic structure that can facilitate or inhibit cell survival by restricting replication fork movement and requires further examination (Goodall et al., 2023; Rudolph et al., 2013; Rudolph et al., 2019).

Overview of DNA replication in bacteria

Initiation

Bacteria with a single circular chromosome regulate replication initiation at a single origin called *oriC* with the help of DnaA protein. Two replication forks are formed at *oriC* and move bidirectionally until they meet at the opposite site where replication terminates. (Bird et al., 1972; Kuempel et al., 1973; Kuempel et al., 1977; Masters and Broda, 1971). Replication in *E. coli* initiates at *oriC* and replisomes proceed bidirectionally into each half of the replichore until they fuse at the arithmetic midpoint creating a chromosomal dimer needed for resolution. Initiation at *oriC* is a DnaA dependent process requiring DnaA binding to specialised box areas flanking *oriC*, in a similar fashion to histone protein coiling in eukaryotes. This coiling allows unwinding of duplex DNA by relieving duplex torsion and facilitates the entry for DnaB helicase to be loaded onto the lagging strand through DnaC-DnaB interactions (Mott and Berger, 2007). Figure 4 shows the step by step processes in order for DnaA to melt duplex DNA and load the replisome (Chodavarapu and Kaguni, 2016). Once the replisome has been assembled, DnaB helicase proceeds to translocate along the lagging strand in a 5' - 3' direction, allowing DNA polymerase III holoenzyme to synthesize the nascent strand continuously on the leading strand and in Okazaki fragments on the lagging strand.

DNA replication in *Escherichia coli* initiates from a single origin on the circular chromosome called *oriC*. *oriC* contains a series of asymmetric 9 base pair DnaA boxes that exhibit high or low affinity binding to the DnaA initiator protein (Messer, 2002; Mott and Berger, 2007). DnaA remains continuously bound to the high affinity sites, while occupancy of the lower affinity sites fluctuates during the cell cycle. Binding of ATP-bound DnaA (the active form) to the low affinity DnaA boxes causes localized unwinding within an AT-rich DNA unwinding element (DUE) adjacent to *oriC*, generating single-stranded DNA regions (Leonard and Grimwade, 2005).

Unwinding of the DUE allows loading of the replicative DnaB helicase in a reaction mediated by DnaA and the helicase loader protein DnaC (Costa et al., 2013). DnaB translocates along the DNA template, forcing the parental strands apart. This enables the DnaG primase access

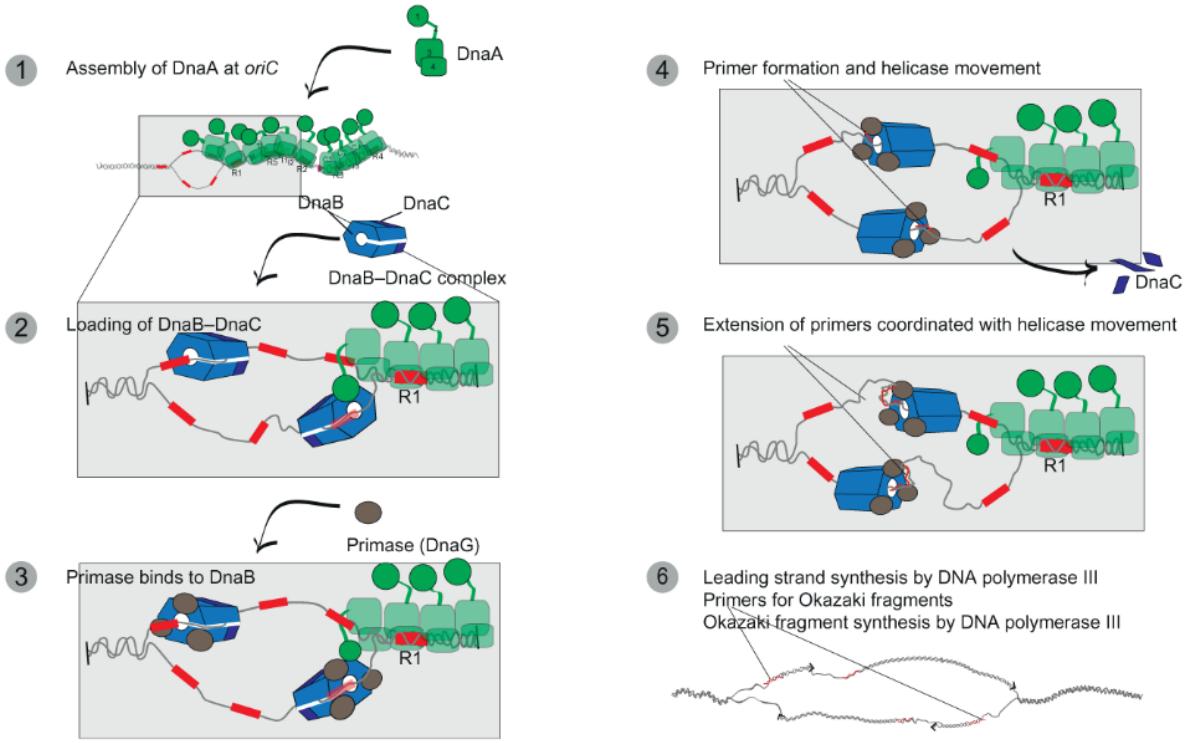

Figure 4: Schematic showing initial stages of replisome assembly at *oriC* via DnaA binding to boxes either side of the region and allowing opening of the DNA duplex and the recruitment of the DnaB-DnaC complex. (Adapted from Chodavarapu et al. 2016)

to synthesize RNA primers that are required for assembly of the replisome machinery (Messer, 2002; Mott and Berger, 2007). Following replisome assembly, the two replication forks move away from *oriC* in opposite directions as elongation commences.

Several regulatory mechanisms operate following initiation to prevent re-initiation events and restrict chromosome duplication to once per cell cycle (Skarstad and Katayama, 2013). These include origin sequestration mediated by SeqA binding to methylated GATC sequences, which prevents DnaA rebinding (Kaguni, 2006; Marinus and Løbner-Olesen, 2014). The regulatory inactivation of DnaA (RIDA) process promotes ATP hydrolysis to convert DnaA to an inactive form (Katayama, 2001). DnaA is also titrated away from *oriC* by high affinity binding sites like *datA* (Hansen et al., 2007; Kitagawa et al., 1996; Skarstad and Katayama, 2013). Finally, *dnaA* gene expression is repressed by DnaA itself through binding to its own promoter region, which is also sequestered by SeqA (Atlung et al., 1985; Campbell and Kleckner, 1990; Hansen and Atlung, 2018; Messer, 2002; Riber and Løbner-Olesen, 2005). Together, these regulatory processes ensure tightly controlled DNA replication initiation in *E. coli*.

E. coli, like most bacteria, replicates its chromosome from a single origin, whereas eukaryotes have thousands of origins spread out over the linear chromosome which is an efficient mecha-

nism to speed up replication when the genome size is orders of magnitudes larger than prokaryotes. Why then do bacteria have their origin restricted to one? This has been a puzzling question for geneticists studying prokaryotic replication, yet due to the fact that forks moving away from *oriC* at a rate between 650 - 1000bp/s, 10-20 fold faster than the speed of transcription (Dennis et al., 2009), and roughly 10 times faster than eukaryotic cells, there may have not been a selection pressure high enough to integrate multiple origins into the genome. Having more than one origin is not uncommon in other single celled organisms, much like the archaea *Haloferax volcanii*, which has 4 origins located around its chromosome. Paradoxically, work by Hawkins and colleagues (Hawkins et al., 2013) showed that deleting all origins in *H. volcanii* actually increased the rate of replication, where it was initially predicted that this would drastically impact the ability of the archaeon to survive. Upon further investigation into the robust growth of these archaea it was reported that the cells lacking functional origin firing can duplicate their chromosome purely from pathways involving homologous recombination (HR), highlighting further evidence for the interplay between replication and recombination (Kreuzer, 2005). Where these processes were once thought as separate systems, we now have the view of their intersection with key proteins and pathways.

Elongation

DNA polymerase III is a high fidelity polymerase able to elongate DNA at a rate of about 1000 bases per second. The two independent replisomes move at roughly the same rate until they approach the termination area where they will fuse to complete replication. As mentioned in the above sections, the speed of replication is orders of magnitude faster than transcription meaning there will often be transcription-replication conflicts. This is not so much an issue in eukaryotic cells, which separates the processes of replication and transcription, likely as an evolutionary mechanisms to avoid these conflicts, however transcription is not switched off altogether during S phase and RNA polymerases (RNAP) can still pose a significant challenge to replicating cells. Early work from Dennis and colleagues have showed varying elongation rates at different *rrn* operon sites during transcription, so the exact location at which replisome may meet RNAP is stochastic by nature (Dennis et al., 2009). RNAPs pose a real threat to the stability of the replisome during replication and these conflicts must be mitigated where possible and resolved efficiently when they do occur, routinely through restarting of the stalled replication fork (Pomerantz and O'Donnell, 2010; Rudolph et al., 2007). A natural by product of transcription is the formation of RNA:DNA hybrid structures, called R-loops, where the complementary

mRNA hybridizes to the DNA template strand and displaces the non-template DNA (Asai and Kogoma, 1994; Gowrishankar et al., 2013; Hong et al., 1995). In Figure 5, we can see how the directionality of the replisome with transcribing RNAPs can stabilise or reduce R-loop formation, depending on if these processes occur head-on or co-directionally, respectively (Lang and Merrikh, 2018).

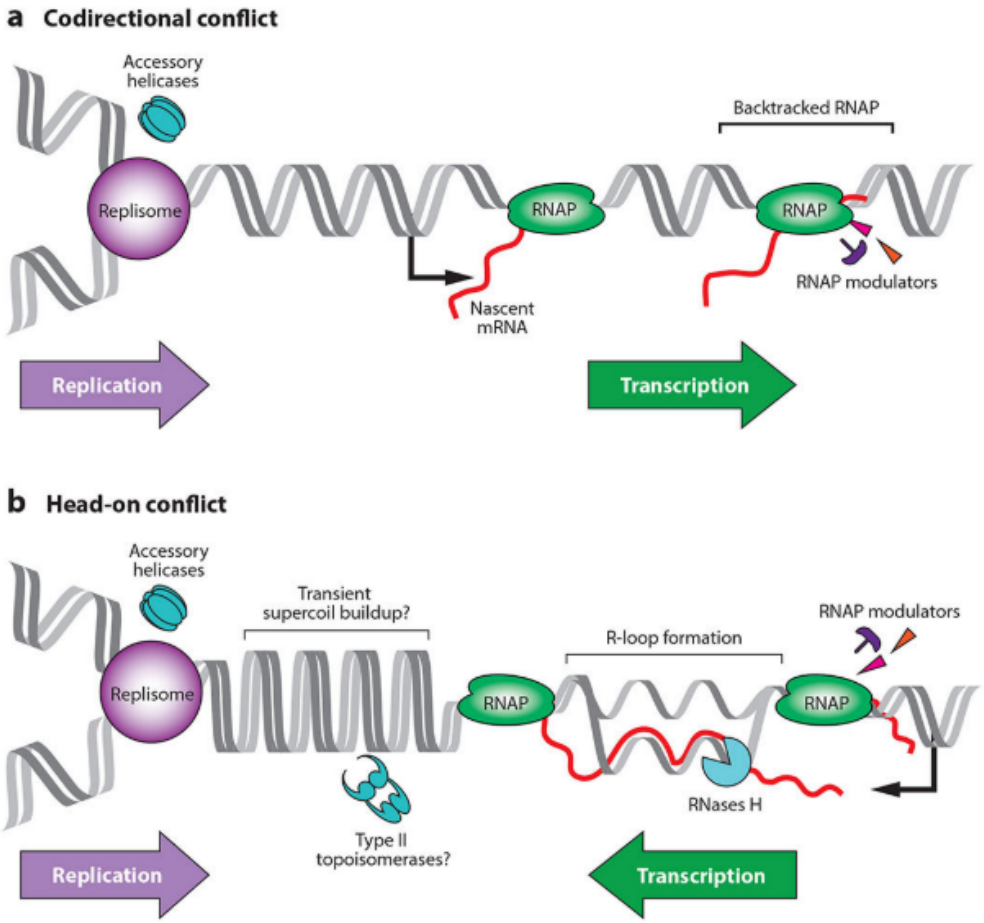

Figure 5: A) Co-directional replication and transcription of replisome and RNAP reduces head-on collisions and associated R-loop formation, RNAP dysregulation, and genomic instability. B) Head-on conflicts cause increased positive supercoiling build up ahead of the replication fork, subsequent negative coiling behind RNAP leading to favourable R-loop conditions and stalling of both the replication fork and RNAP. (Taken from from (Lang and Merrikh, 2018))

Homologous Recombination in Bacteria and Types of DNA Damage

DNA integrity is vital for faithful transmission of genetic information. However, DNA is constantly subject to damage from both endogenous and exogenous sources. These lesions, if left unrepaired, can impede replication and transcription - potentially blocking these essential cellular processes. We will discuss two major forms of DNA damage: single-stranded breaks and double-stranded breaks. The intricate repair mechanisms that have evolved to resolve these le-

sions are complex and research is constantly ongoing to further elucidate mechanisms of DNA repair. During the termination stage of DNA replication, where replisome dissociation must be coordinated with repair of any remaining discontinuities, the repair pathways mentioned here are critical for completing replication and safeguarding the genome. The interplay between DNA repair and replication is critical in *E. coli*. Replication-blocking lesions need to be repaired efficiently to allow replication restart and avoid extended fork stalling. Meanwhile, certain types of repair can also be triggered by replication errors and fork collapse (Dimude et al., 2016; Midgley-Smith et al., 2018; Rudolph et al., 2013).

Base excision repair (BER) acts on small base lesions like oxidative damage that cause relatively minor helix distortion. These lesions can transiently block replicative polymerases, leading to misincorporations and mutations. BER helps counter replication errors by fixing such lesions in a post-replicative manner. The DNA glycosylases that initiate BER tend to have overlapping specificities, providing redundancy. BER can rapidly repair lesions outside of replication via the short-patch subpathway.

More bulky lesions trigger nucleotide excision repair (NER), which can more extensively stall replication forks. NER repairs these bulky adducts and UV-induced lesions by excising a 12-13 nucleotide fragment surrounding the damage (Newton et al., 2012). NER links with replication in two ways; Transcription-coupled NER (TC-NER) is triggered when RNA polymerase stalls at a lesion and recruits repair factors like Mfd. TC-NER handles lesions on the transcribed strand, and global genomic NER (GG-NER) scans the entire genome and handles non-transcribed lesions. It is initiated by UvrAB complexes and present at higher levels in rapidly replicating cells. The UvrABC nuclease complex plays a key role in detecting and verifying lesions before excision. UvrD helicase unwinds the DNA duplex allowing repair synthesis by DNA polymerase I. Generally, NER activity helps avoid replication fork collapse at bulky lesions (Atkinson et al., 2009), which eventually triggers replisome dissociation and duplex resection by RecBCD at double strand ends (Courcelle, 2005). This generates substrates for recombinational repair (RR) mediated by RecA and the RecFOR proteins. RR uses homologous recombination to restart collapsed forks and fill in single-stranded gaps. The precision of recombination maintains replication fidelity (Morimatsu and Kowalczykowski, 2003; Sakai and Cox, 2009).

DNA damage generates the ssDNA signal that activates RecA to RecA* (Kovačič et al., 2013). RecA* stimulates cleavage of the LexA repressor, inducing expression of SOS genes (Rehrauer et al., 1996) . This allows cells to repair damage, but at the cost of increased mutation. As damage

is repaired, LexA re-accumulates and shuts off the response. The interaction between RecA* and LexA is thus the key event that controls the switch between repressed and induced states of the SOS response. The lexA gene encodes the repressor that keeps SOS genes off, while the recA gene encodes the sensor that detects DNA damage and induces the response (Rehrauer et al., 1996).

Arguably, the most cytotoxic form of DNA damage are double strand breaks (DSB) which result in a fully detached section of the DNA that requires recombination protein complexes for repair. When replication forks approach DBSs, the replisome always runs off the DNA which will require fork restart at the location of the break for successful chromosome duplication (White et al., 2018). DSBs sever both strands of the double helix, potentially deleting large segments of genetic information if not rapidly repaired and in bacteria, DSBs are primarily repaired by two pathways: homologous recombination (HR) and non-homologous end joining (NHEJ).

HR utilizes the information in an undamaged homologous sequence to achieve accurate repair. In *E. coli*, HR is initiated by the RecBCD complex. Upon binding a blunt DSB end, RecBCD unwinds and simultaneously degrades the DNA duplex in an ATP-dependent manner [Courcelle (2005); Sakai and Cox (2009);]. it is worth noting that RecBCD can also process DNA ends with short 3' or 5' overhangs, as long as they are within the binding range of the RecBCD complex (Lenhart et al., 2014). Degradation continues until RecBCD encounters a *chi* sequence in the proper orientation, after which RecD nuclease disengages its 3' - 5' activity, altering the RecBCD complex to form 3' ssDNA filaments on the same strand as the *chi* site (Amarh et al., 2018). RecBCD then continues to generate the 3' filament whereby RecFOR can now loads RecA onto the emerging 3' single-stranded overhangs it generates (Morimatsu and Kowalczykowski, 2003). RecA polymerizes into a nucleoprotein filament that facilitates homology search and invasion of the intact sister chromatid. Next, RecA catalyzes strand exchange, allowing the missing sequence to be reconstituted by DNA polymerase using the homologous sequence as a template. Finally, ligation restores continuity on both strands. Thus, HR provides error-free repair by relying on an undamaged homologous template (Lenhart et al., 2014).

While most homologous recombination in bacteria relies on RecA, an alternative RecA-independent pathway exists that is important for repairing post-replication gaps. Recent work by Jain and colleagues demonstrated that this RecA-independent recombination depends largely on the conserved ATPase activity of RarA (Jain et al., 2021a, 2021b). Intermolecular recombination rates between plasmids sharing short repeated sequences (104-411 bp) were

significantly reduced in *ΔrecA ΔrarA* double mutants compared to *ΔrecA* alone. The ATPase activity of RarA was required, indicating it catalyzes a strand exchange reaction analogous to RecA. Moreover, deleting *rarA* ameliorated growth defects of *xerD* mutants, which accumulate chromosomal dimers. This suggests RarA promotes crossovers that must be resolved by XerCD. Although RarA contributes little to recombination when RecA is present, it facilitates a parallel pathway prominent during repair of short patch gaps (Jain et al., 2021a, 2021b).

Exonucleases also suppress an alternative recombination pathway, as deleting *recJ* or *exoI* in *ΔrecA ΔrarA* cells increased recombination rates. Accessory factors like RecF, RecO and translesion polymerases may additionally contribute. While vestigial recombination remains when both RecA and RarA are absent, this work significantly advances understanding of RecA-independent gap repair. RarA is now firmly positioned as mediating template switching to mitigate lesions skipped by the replisome. Ongoing efforts to elucidate its mechanism will provide further insight into the interconnected pathways preserving genomic stability during DNA replication (Jain et al., 2021a, 2021b).

Fork Restart

Replication fork restart is essential for completing DNA replication and maintaining genome stability in bacteria. When replication forks stall due to DNA damage or other obstacles, they must be restarted in order for replication to finish properly. Otherwise, incomplete replication can lead to chromosome breaks and rearrangements. Research over the past two decades has revealed key pathways and proteins involved in fork restart.

The major causes of fork stalling necessitating restart have been identified. DNA damage, especially nicks and lesions, frequently block replication forks (Cox, 2001; Cox et al., 2000; Michel et al., 2018). Fork stalling also occurs when replication conflicts with transcription complexes on the DNA template, particularly with head-on collisions of the replisome and RNA polymerase (Rudolph et al., 2007). Finally, fork stalling can occur stochastically due to replisome disassociation from the DNA template, estimated to happen nearly once per cell cycle (Windgassen et al., 2018). Figure 6 shows a model for Replication Restart pathways. There are PriA-dependent and PriA-independent pathways that all lead to DnaB loading onto ssDNA (Sandler et al., 2021). The PriA-dependent pathways involve PriA, PriB, and DnaT (Mahdi et al., 2012). There are three PriA pathways proposed: PriA-PriB1, PriA-PriB2, and PriA-PriC. PriA-PriB1 and PriA-PriC require helicase activity but PriA-PriB2 does not. PriA-PriB2 and PriC provide ssDNA similarly by re-

modeling SSB. SSB tetramers may be present on the ssDNA. RecG helicase also plays a role here, helping prevent unnecessary replisome assembly events mediated by PriA and PriB that can cause pathological over-replication (Mahdi et al., 2012; Rudolph et al., 2013). RNA polymerase mutations enhance *ΔpriB* suppression of *recG* by reducing replication-transcription conflicts. This avoids exposing single-stranded DNA substrates for unnecessary recombination.

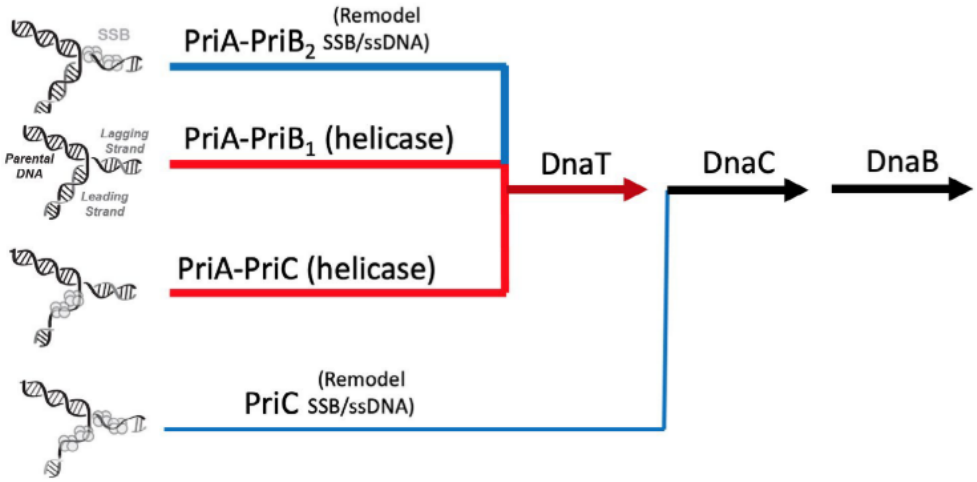

Figure 6: Refined model of replication restart pathways highlights multiple PriA-dependent and PriA-independent mechanisms convening at DnaB helicase loading. Potential substrates are shown on the left, with parental DNA in black and new strands in grey. (From Sandler et al. (2021))

When forks stall, they can undergo structural changes influencing the restart mechanism. Fork regression or reversal can occur, in which the fork "rewinds" itself into a chicken foot structure with a Holliday junction (Cox, 2001; Heller and Marians, 2005). Alternatively, single-stranded nicks at the fork can cause fork breakage and double strand breaks (Michel et al., 2018). The structure of the stalled fork affects which proteins are required for restart.

Multiple pathways catalyze fork restart in bacteria. The PriA protein recognizes and remodels stalled forks and orchestrates helicase loading for restart (Heller and Marians, 2005; Windgassen et al., 2018). Two main pathways using PriA have been defined. The PriA/PriB pathway utilizes PriB and DnaT to assist PriA in reloading the replicative DnaB helicase (Heller and Marians, 2005; Windgassen et al., 2018). This pathway is efficient on forks with an intact leading strand (Heller and Marians, 2005). The alternative PriA/PriC pathway instead utilizes PriC to load DnaB, likely also requiring DnaT (Heller and Marians, 2005; Windgassen et al., 2018). This pathway preferentially acts on forks with gaps in the leading strand (Heller and Marians, 2005). An additional PriC/Rep pathway was more recently discovered that utilizes the Rep helicase to assist PriC in loading DnaB independently of PriA, but this has only been

reconstituted *in vitro* (Windgassen et al., 2018).

If stalled forks have undergone breakage, recombination pathways are required before PriA-mediated restart. RecA, RecBCD, and RuvABC recombinases process and repair broken forks to generate structures amenable for PriA loading of DnaB (Michel et al., 2018; Windgassen et al., 2018). In contrast, direct restart of stalled but intact forks via PriA pathways only requires reloading of the replicative helicase DnaB without need for recombination first (Michel et al., 2018).

Given that forks often become stalled and require repair and/or restart to fully replicate the chromosome, there is strong evidence that this repair is recombination dependent. The Cox lab have shown roughly half of cells growing under normal conditions are dead when RecA is inactive, the facilitator of strand invasion that is loaded onto DNA paired with SSB. This effect is even worse in cells with a *priA* null mutation, suggesting that replication fork restart is a normal housekeeping function of bacterial cells (Cox et al., 2000).

Replication forks can encounter various obstacles that can stall or block their progression, such as protein-DNA complexes, G4 quadruplex DNA or DNA lesions (Bianco, 2020; Lang and Merrikh, 2018; Linke et al., 2021; Marians, 2018). These obstacles can occur randomly in different chromosomal locations during each round of genome duplication. However, some bacteria have evolved mechanisms to specifically arrest replication forks at certain sites (Hizume and Araki, 2019; Hyrien, 2000; Labib and Hodgson, 2007). For example, in plasmid systems like R6K, replication can be either unidirectional or bidirectional depending on the initiation mechanism (Abhyankar et al., 2003; De Graaff et al., 1978; Rakowski and Filutowicz, 2013), but both forks are blocked by a specific terminus site (Crosa et al., 1976). In the R1 plasmid, a distinct termination system enforces unidirectional replication by blocking the counter-clockwise fork with two *ter* sites located between the minimal origin *oriR* and the start site for leading strand replication (Hill et al., 1988; Krabbe et al., 1997). Replication fork arrest can also result from direct fork breakage by nucleases at specific sites that block bacterial chromosome replication, such as replication terminator sequences (Michel et al., 1997).

When replication forks encounter obstacles such as DNA lesions, fork stalling and collapse can occur, requiring fork restart mediated by homologous recombination. Processing of stalled forks by RecBCD generates double-stranded DNA ends, which are recombinogenic substrates (Michel et al., 2018). RecA polymerizes onto the 3' single-stranded tails and catalyzes strand invasion into homologous duplex DNA, forming a displacement loop (D-loop). D-loops represent

an intermediate in repair and restart of collapsed forks but can also be initiating structures for pathological over-replication in certain mutants.

Interplay between Replication and Recombination

Early research into mechanisms of replication and recombination were segregated for many years, until evidence gradually mounted for an interlay between the once thought separate processes (Briggs et al., 2003; Kreuzer, 2005). Early work from Kogoma and co-workers (Asai and Kogoma, 1994; Kogoma, 1997) showed how replication can initiate at a number of DNA structures using HR or transcription as their driver. They highlight how initiation in these pathways is dependent on the recombinational protein RecA, PriA restart protein and others (DnaB, DnaC and DnaG primase) which all facilitate loading of a nascent replisome onto a site of duplex opening. They termed this form of replication stable DNA replication (SDR) and subdivided the type of SDR based on whether or not initiation began at an R-loop or D-loop and the conditions needed to allow this to take place. Both forms of SDR open the DNA duplex to allow strand invasion, with inducible SDR (iSDR) defined by a single stranded DNA (ssDNA) end invading the duplex forming a D-loop, and constitutive SDR (cSDR) defined as RNA transcript invasion forming an RNA-DNA hybrid. These forms of replication were also categorised as recombination dependent replication (RDR) (Asai and Kogoma, 1994; Briggs et al., 2003; Kogoma, 1997; Kreuzer, 2005).

D-loops, or displacement loops, can also be the site for replication initiation. Homologous recombination is a process that allows the repair of certain DNA damage and has been shown to restart stalled replication forks. If replicating DNA encounters a lesion or single strand gap (SSG) there is a chance there will be a double strand end (DSE) formed. DSEs are known substrates for RecBCD, which possesses nuclease activity and degrade DSE at a rapid rate. When RecBCD encounters a *chi* site there is a conformation change that switches off the 3'-5' nuclease activity, but enhances the 5'-3' nuclease activity (Dimude et al., 2018a; Hamilton et al., 2019; White et al., 2018). After a while there is a 3' overhang which can be targeted by SSB and eventually RecA loading. Any ssDNA with an exposed 3'OH end is canonical substrate for many enzymes which utilize recombination. After RecA is bound to the ssDNA, the recombinase facilitates strand invasion of a homologous chromosome, displacing one arm of the sister chromosome and integrating itself into the duplex, forming a D-loop Dimude et al. (2015). This process occurs as part of the repair pathway in *E. coli* following not only DNA damage, but also stalling of a replication fork. Stalled forks need to be restarted in order to maintain cell survival and successful genome

duplication, discussed later on in the introduction. From Figure 7 we can see an illustration of the proposed model for how the RecBCD enzyme complex controls DNA synthesis at double-strand breaks. When RecBCD is present, it coordinates the invasion of both broken DNA ends into the sister chromosome, generating two converging replication forks that replace lost genetic information. If one fork establishes before the other and reverses the second fork, RecBCD can degrade the extruded end to restore the fork. Without RecD, either uncoordinated end invasion occurs or coordinated invasion is followed by fork reversal that RecBC cannot resolve. In the absence of RecD, RuvABC Holliday junction resolution converts coordinated to uncoordinated invasion. Either uncoordinated scenario allows palindromic re-cleavage, chromosome amplification, and loss of viability (White et al., 2018).

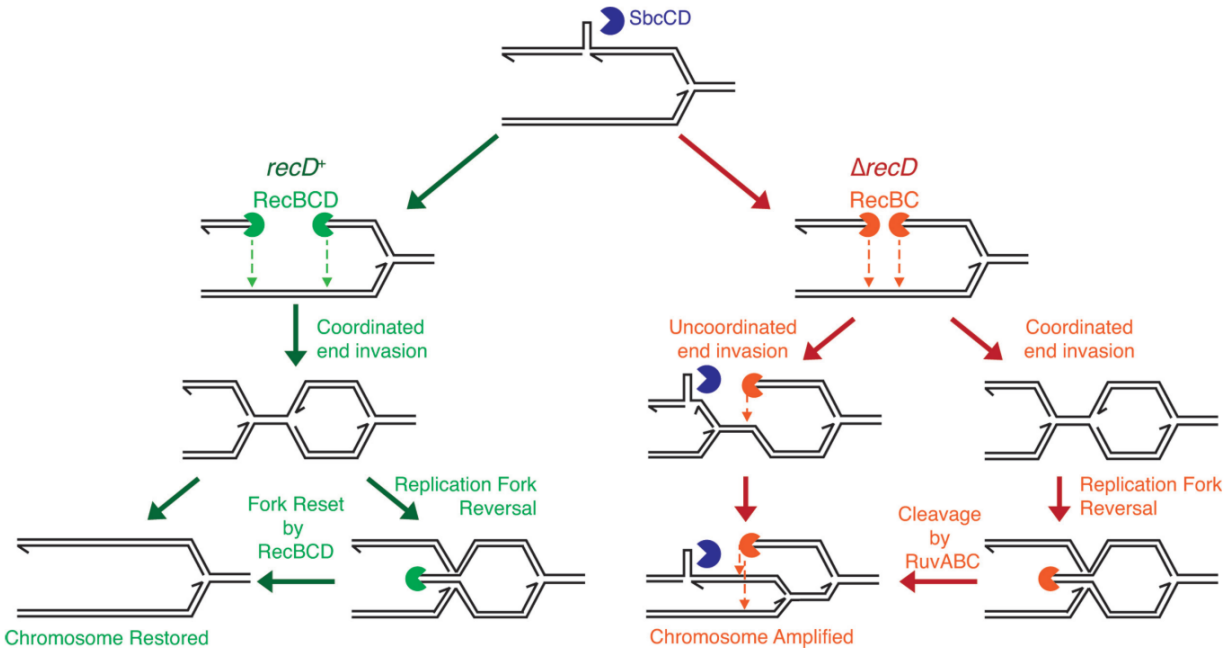

Figure 7: Model and pathways of RecBCD mediated DNA repair of lesions and double strand breaks (Taken from (White et al., 2018))

In prokaryotes, such as *Escherichia coli*, errors occurring during DNA replication that reduce genomic stability may affect the growth rates and survival of cells through inappropriate DNA metabolism leading to accumulation of DNA structures which interfere with the movement of the replisome during replication (Drolet and Brochu, 2019). Errors that occur during segregation of sister chromosomes at the end of DNA duplication may also cause unwanted stress and genomic instability during the final stages of replication (Bigot et al., 2005; Midgley-Smith et al., 2019). In eukaryotes, this genomic instability could potentially lead to genetic disorders or even cancer (Negrini et al., 2010; Wei Dai, 2014). It has been highlighted that a number of human disorders stem from a genetic basis in which one, or a number, of mutations is responsi-

ble for the development of pathology. However, oncogenesis in humans is more nuanced than previously thought, with recent research suggesting that issues stemming from errors in DNA replication that lead to the accumulation of mutations or DNA intermediate structures are a key driver in cancer development (Vassilev and DePamphilis, 2017). This genomic instability can be thought of as an umbrella term which classifies errors occurring during DNA replication elongation and termination, errors in mismatch repair, DNA damage and double strand break repair (DSBR) which require homologous recombination processes (Bishop and Schiestl, 2002), as well as DNA-RNA hybrid structures called R-loops (Crossley et al., 2019). Thus, the need to reduce genomic errors and maintain stability during all stages of DNA replication is in the interest of all organisms, single or multicellular (Cox et al., 2000; Hyrien, 2000). Any genomic system(s) that can be identified which have evolved to reduce the likelihood of instability is therefore of paramount importance and will add crucial details to our understanding how one cell becomes two.

Interplay of R-loops and G-quadruplexes

In wild type cells, the *rnhA* gene encodes RNase HI and removes RNA:DNA complexes, known as R-loops, which arise from negative supercoiling and duplex opening by an invading RNA transcript. By inactivating *rnhA* cells can initiate replication at accumulated R-loops. Although there has been difficulty in identifying these *oriK* sites, it was shown that *ter* sites are 'hot spots' for recombination and the termination area as a whole is a recombinational hotspot (Asai and Kogoma, 1994; Horiuchi et al., 1994). Building on from their work, data has accumulated that cells lacking *rnhA* show over-replication inside the termination area, ruled out as being due to a cryptic prophage origin, supporting the idea of SDR within the termination area (Maduike et al., 2014). R-loops can be the site of initiation when the cell has entered the SOS response and *oriC* firing has stopped (Kogoma, 1997), however this process is dependent on transcription and is inhibited by rifampicin. These studies also highlight another recombination protein that has the ability to remove R-loops and is inversely correlated with SDR, the DNA translocase RecG.

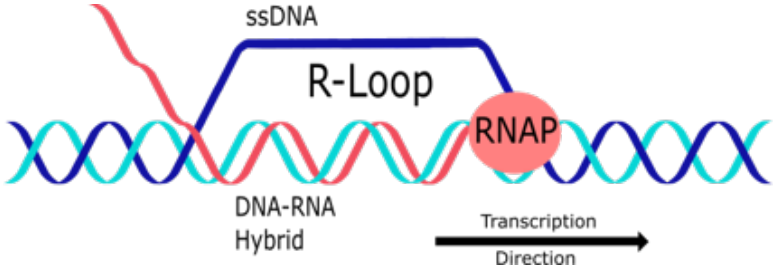

Figure 8: Schematic of R-loop formation and RNA hybridisation to DNA template. Rendered with Inkscape

Forks meeting RNAP barriers will become arrested if the barrier isn't removed, halting the replication process leading to downstream repair pathways that can increase genomic instability. Rupp and Howard-Flanders (1968) proposed that when a fork encounters a lesion it simply skips over it, leaving gaps of unsynthesized DNA, that are filled by RecA mediated homologous recombination. More recent research has shown that lagging strand lesions are more tolerable than template strand lesions, with the lagging strand able to simply bypass the lesion. However, the situation is more serious in either of the template strands as forks unwinding the DNA will generate a DSE when encountering a template lesion. This lesion will collapse the replication fork, disassemble the replisome and require HR to restart the fork. This process has been shown to be dependent on RecBCD and RecA, as previously described with the process of D-loop formation that can then be targeted by the restart protein PriA (Rudolph et al., 2007).

Experimental work from Dimude and colleagues (Dimude et al., 2015) have shown that there is

over-replication in cells lacking RNase HI, and this is restricted to certain sites, namely inside the termination area. From recent findings suggesting that over-replication occurs from accumulation of 3' flaps, the authors wanted to find out if the *rnhA* over-replication occurs from the same mechanism. Although RecG and RNase HI are both able to remove RNA from R-loop complexes *in vitro*, it appears that the over-replication following their inactivation are through different means. Following the accumulation of R-loops, in cells lacking RNase HI, the replication restart protein PriA is able to target the branched structure and assist in loading DnaB helicase onto the lagging strand followed by the remaining replisome proteins. From Figure 8, we can see how over-replication in *rnhA* single mutants are dispersed throughout the chromosome, initiated at R-loop structures.

To maintain genomic stability, transcription and replication share a high degree of co-directionality in the *E. coli* chromosome to minimise collisions between replisomes and RNA polymerases. If this type of collision takes place, fork stalling can occur which can severely impede growth and interfere with genomic stability and cell survival. For prokaryotes this means growth defects and possible inviability, whereas for eukaryotes this can mean an increased risk of cancer or neurodegenerative disease (Brambati et al., 2015). Suppressor mutations often occur in mutants where collisions are likely to take place, such as with ectopic origins or in cells which depend on over-replication within the termination area to duplicate the chromosome (Dimude et al., 2015, 2018b; Rudolph et al., 2007). In *E. coli*, more than 50% of all genes are oriented on the leading strand template meaning they are co-directional with replication. Generally, highly important genes for exponential growth are positioned on the leading strand, which allow fast and efficient gene expression whilst DNA is being replicated. For genes present on the lagging strand, this poses a problem for cells in exponential growth phases, how do they bypass RNAPs in order to continue duplicating the chromosome? The slow growth, and sometimes inviable, phenotype is well documented in cells which have increased collisions between these two processes, resulting in not only reduced replication and gene expression, but also mutagenesis in the affected gene (Schroeder et al., 2016). It has been suggested that following collisions between the replisome and RNAP transcribing along the lagging strand, an R-loop is able to form behind the RNAP, which further exacerbates the pathogenesis of stalling a fork (Brambati et al., 2015; Lang et al., 2017). Recombination proteins such as RecG have been shown to unwind R-loops which minimises the conflicts between replication and transcription, this coupled with its ability to reduce over-replication of the termination area granted RecG the title of guardian of the bacterial genome, yet more research into its precise function is still

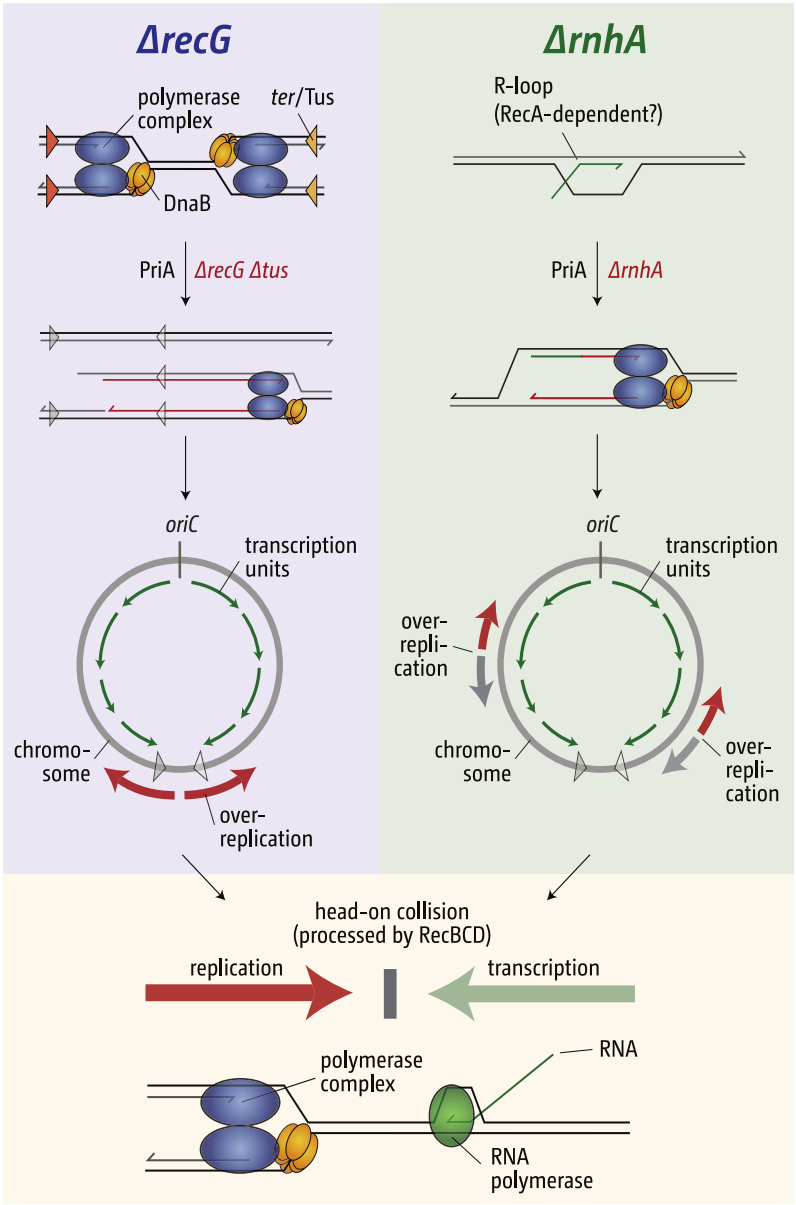

Figure 9: Schematic showing replisome colliding head on with RNA polymerase. This scenario is particularly challenging to the cell when the replication fork trap has been inactivated as replication forks exiting the terminus region will continue towards *oriC* in which they will meet RNAP during transcription. Adapted from Dimude et al. (2015)

needed even after 25 years of study (Lloyd and Rudolph, 2016; Rudolph et al., 2010a).

Recent genome-wide mapping of R-loops in human cells via DRIP-seq and bisulphite footprinting reveals new insights into the prevalence and characteristics of these structures *in vivo* (Malig et al., 2020). R-loops were found to be abundant, with strong concordance observed between the two techniques. The majority of mapped R-loops ranged from 200 to 500 base pairs in length, although some exceeded 1 kilobase (Malig et al., 2020). Notably, R-loops were concentrated in discrete clusters across the genome rather than distributed randomly. These clustered R-loops appeared to pile up into larger zones of enrichment, with boundaries not readily predicted by DNA sequence alone. Intriguingly, R-loops derived from pre-mRNA were found to frequently span splice junctions. The work by Malig and colleagues demonstrate the nonrandom distribution of R-loops into clustered piles with sizes larger than individual transcripts, highlighting these structures are not necessarily formed purely from stochastic means.

The R-loop structure is likely stabilized due the topological changes occurring at the DNA during collisions between replisomes and RNAP. As RNAP facilitate duplex opening to read the template DNA, this results in positive supercoiling ahead of transcription and negative supercoiling behind the RNAP (Brochu et al., 2023; Drolet and Brochu, 2019). These topological changes also affect rates of R-loop formation as head on collisions between replisomes and RNAP increase the positive coiling, thereby also increasing the negative coiling behind the RNAP and allowing for an easier R-loop formation.

At the DNA between the two complexes there is positive supercoiling, increasing torsional tension and making it difficult for helicases to further unwind DNA the closer they get to one another. This also has the effect of introducing negative supercoiling behind RNAP, allowing easier access for transcript invasion of the duplex. (Lang et al., 2017). Removing R-loops and restarting the replisome is of great importance for cell survival. At the heart of this issue is trying to remove the R-loop blocking the path of the replisome and alleviating the torsional stress imposed by the collision. One study (Hamperl et al., 2017) suggests that replisomes act as regulators for the formation of R-loops *in vivo*, with head on collisions promoting R-loop formation and co-directional collisions ameliorating their formation. In *E. coli*, the main nuclease to degrade RNA from R-loop is RNase HI, and as described above, cells lacking RNase HI show increased levels of SDR. It is likely that replisome collisions with RNAPs occur *in vivo* during every round of replication. When genes are oriented co-directionally with the the replication fork, the cells seems able to cope with these two processes using the same template DNA. One study showed

that the replisome is able to bypass RNAP while remaining attached to the DNA, causing RNAP to dissociate and the replication fork to continue without collapsing. The nuance being, that the researches found that for both the T7 bacteriophage and *E. coli* RNAP the leading strand essentially stops synthesis and uses the mRNA:DNA hybrid to restart synthesis (Pomerantz and O'Donnell, 2008). This is supported by data showing that a point mutation in the *rpoB* gene at position 35 (rpo^{*35} referred to hereafter as rpo^{*}) will destabilize the ternary structure of RNAP and allow replisomes to move past these transcription barriers without becoming stalled. This mutation has been used to clarify if cells are inviable due to collisions with highly transcribed *rrn* operons in cells with ectopic origins *oriX* and *oriZ* (Dimude et al., 2015; Ivanova et al., 2015; Syeda et al., 2020).

The fact that most researchers agree head-on transcription-replication conflicts are detrimental to the cell has posed the question of why bacteria haven't evolved to align all their genes with replication? Surely minimizing all conflicts by arranging the genome to fit replication-transcription co-directionality would be of benefit to the cell and result in increased survival? Data from Merrikh and colleagues (Merrikh, 2017; Merrikh and Merrikh, 2018; Merrikh et al., 2016) have shown that head-on transcription-replication conflicts increase the incidence of mutagenesis that can result in adaptive mutations and actually increase cell survival. This comes from the fact that many stress induced genes are oriented on the lagging strand so that head on collisions with transcription machinery can actually increase survival rates through gene specific mutagenesis (Schroeder et al., 2016, 2020). Thus, although co-directionality is favoured overall for highly transcribed genes there may be a requirement for the cell to maintain some level of lagging strand genes which can be called upon in times of environmental or genetic stress to increase the probability of survival by increased mutation rates.

Taken together it seems that conflicts with transcription are encountered under normal conditions and pathways have been identified that reduce the consequences of such conflicts. It has been shown that accessory replicative helicases, which do not function through recombination, assist in restarting replication as the replisome encounters RNAP. Rep and UvrD are the most extensively studied helicases that translocate ssDNA in the 3'-5' direction and promote replication fork progression through nucleoprotein barriers, of which RNAP is the most common given that transcription and replication occur simultaneously (Epshtein et al., 2014). Although these proteins share homology and a general function they function differently and localise at the replisome through different interactions. In fact, UvrD is associated with the replisome, not through any binding activity to 'hitch a ride' but it is localised there due to its function in

removing barriers and assisting the progression of the replisome (Wollman J. et al., 2023). Rep directly binds to DnaB helicase of the stalled replisome, whereas UvrD does not bind to any subunit of the replisome and seems to be localised at the stalled replisome solely due to its function in overcoming the RNAP barrier. UvrD also interacts with RNAP and has been implicated in pulling stalled RNAP backwards to allow DNA repair machinery to gain access to damaged DNA (Epshtein et al., 2014). DinG also contributes in this context although its function is unclear, while RecG plays a role in reducing R-loop accumulation by unwinding the structure to displace the RNA (Hawkins et al., 2019). It is likely, therefore, that removal of protein-protein barriers is an essential housekeeping function of cells to successfully complete replication.

It has been reported that replisomes can remain stably bound at replication barriers as long as the fork does not run into a lesion or gap on the leading strand. When this happens the replisome will likely become disassembled in order to repair the lesion, requiring fork arrest and repair by recombination (Bianco, 2015). One method replisomes may remain stably bound to forks when encountering a lesion or barrier which would normally result in fork collapse, is the reversal of replication forks to form a four strand junction resembling a Holliday junction, sometimes referred to as a 'chicken foot' structure (McGlynn and Lloyd, 2001). This fork reversal has been reported to be catalysed by RecG and RuvAB. However, RecG appears to only play a role in fork reversal when DNA damage is the cause of the fork arrest to form HJ which can resolve the damage (Singleton et al., 2001). These reversed forks can be processed by a number of different recombination pathways in order to restart the replication fork. RuvABC resolvase is required to resolve the structure, which depends on the nuclease activity of RuvC. RuvAB is capable of branch migration and extending the structure after it has been created. RecG translocates along the leading and lagging strand simultaneously and shows affinity for three and four stranded DNA structures (McGlynn and Lloyd, 1999). RecG has three domains and the one implicated for fork reversal is the wedge domain (Briggs et al., 2003). RecG unwinds the nascent strands and reanneals the parent strand to bring the lesion back into the duplex, which can then be repaired by NER systems. The full details of how RecG functions at stalled replication forks has not been fully realised and there is some debate about its use to reverse stalled forks.

The discovery of constitutive stable DNA replication (cSDR) operating independently of canonical DnaA-oriC initiation in RNase HI mutants represented a seminal advance in understanding alternative bacterial replication modes (Kogoma, 1997). However, significant questions persisted about the biological relevance and genomic consequences of this intriguing R-loop-dependent replication pathway. Recent elegant work from the Drolet lab has re-ignited interest

in this phenomenon by demonstrating that unregulated R-loop formation in Escherichia coli lacking topoisomerase I and/or III elicits extensive over-replication specifically within the terminus region, triggering dramatic chromosome segregation defects (Brochu et al., 2018, 2023). This stimulates replication stress that inhibits growth. Delineating the intricate mechanisms connecting R-loop metabolism and replication control in topoisomerase-deficient bacteria provides fundamental new insights into genome maintenance while spotlighting the critical roles of topological regulation in suppressing genomic instability.

Prior work hypothesized that R-loop structures could template unscheduled replication initiation specifically within the termination area, thereby directly causing over-replication of that region and reducing chromosomal replication complexity (CRC) (Kuzminov, 2016). The pivotal discoveries by the Drolet lab has shown that strong *oriK* hotspots supporting R-loop-dependent cSDR replication do not actually map within the over-replicated terminus region in topoisomerase mutants (Brochu et al., 2018, 2023). Deleting known replication fork trap genes at *ter* peaks or inverting *ter* sites failed to abolish over-replication. This argues against models where R-loops directly template replication initiation specifically in the termination area, contrasting prior hypotheses (Kuzminov, 2016). Instead, current evidence supports an elegant model whereby unscheduled replication events fired stochastically by R-loops scattered across the chromosome increase the number of forks traveling to the termination area. This exacerbates fork convergence problems at Tus termination sites in the absence of topoisomerase decatenase activity, culminating in dramatic DNA amplification resembling uncontrolled fork fusion events (Rudolph et al., 2013). R-loop formation genome-wide therefore indirectly promotes replication stress at *ter* by overloading the termination system.

But how might R-loops instigate such widespread inappropriate replication initiation in the first place? While mechanistic details remain sparse, (Brochu et al., 2023) provide clues implicating transcription-associated negative supercoiling, which promotes backtracking to expose 3' RNA ends at R-loops to prime replication. This is appealing, as low-efficiency backtracking could rationalize the observed prevalence of weak stochastic initiation events. However, R-loop-dependent replication could also potentially restart stalled forks at transcribed genes (Lang et al., 2017). Clearly, delineating precise molecular pathways connecting R-loops to replication initiation represents an outstanding goal for the field. It will be important to determine whether hotspots exist at highly expressed loci, and to elucidate signaling cascades specifically activating replication at R-loops versus other aberrant structures.

While R-loops clearly drive over-replication indirectly by overloading Ter, lack of topoisomerase III activity to decatenate interlinked chromosomes appears largely responsible for the aberrant DNA amplification observed in the termination area in these mutants. Strikingly, topoisomerase IV overexpression mitigated massive chromosome segregation defects without reducing R-loop levels, strongly implicating topological causes underlying R-loop-associated toxicity (Brochu et al., 2023). This potentially reflects conserved roles for TopoIII in modulating replication fork topology. Nonetheless, implications emerge for R-loop-triggered genomic instability through downstream replication impacts—unregulated fork progression and collisions with transcription in the wrong orientation can generate hazardous double-strand breaks and loss of viability (Dimude et al., 2015). Thus, R-loop-induced replication stress likely contributes to the profound phenotypes of topoisomerase-defective bacteria. However, precisely mapping sites of replication initiation and delineating why certain loci are favored will be key to evaluate this idea further.

While this research significantly advances understanding of links between R-loop metabolism and replication control, limitations remain. Most pressingly, precisely pinpointing origins of R-loop-dependent replication initiation across the full chromosome at high resolution could reveal informative hotspot patterns or associations with highly expressed genes. Current approaches also need extending to elucidate specific pathways promoting replication initiation at R-loops versus other problematic DNA structures (Kumar and Remus, 2023). Additionally, deciphering the surprisingly complex interplay between RNase HI and topoisomerases in avoiding R-loop-mediated negative supercoiling presents an ongoing challenge. Nevertheless, this work convincingly cements unconstrained R-loop-associated replication as a major driver of genomic instability in topoisomerase-defective bacteria, while spotlighting the critical importance of topological homeostasis in averting such problems.

These discoveries provide evolutionary context by implicating topoisomerases as pivotal regulators suppressing toxic R-loop consequences on replication and genome integrity. While further dissection is needed, evidence supports conserved type IA topo roles in R-loop suppression from bacteria to humans (Zhang et al., 2023). Moreover, connections likely exist with growing links between R-loop-associated replication stress and human disease (Crossley et al., 2019). This progress considerably stimulates the burgeoning field of R-loop biology. A key outstanding question is whether R-loop-dependent replication plays adaptive physiological roles in normal cells, or strictly causes genome instability when unregulated.

When interpreted through the lens of DNA replication termination mechanisms, the identification of R-loop-stimulated over-replication in topoisomerase-defective bacteria provides tantalizing molecular insights. Tus-mediated polar fork traps have elegantly evolved to ensure replisome collisions and termination occur within a defined Ter region (Duggin and Bell, 2009). However, this system appears overwhelmed when unscheduled replication doubles fork numbers genome-wide, preventing proper chromosome segregation. Topoisomerase deficiencies likely compound such problems by hampering decatenation. This paints a picture of multiple intricately coordinated layers of regulation cooperating to successfully conclude replication while preserving genome integrity. When unchecked R-loop-associated replication subverts this, Tus-*ter* traps become inverted from protective mechanisms into drivers of genomic catastrophe. Elucidating how cells dynamically modulate R-loop and topoisomerase levels during growth transitions or stress to avert such outcomes is an outstanding future challenge.

Recent work from the Drolet group significantly propels understanding of the complex intersection between replication regulation, R-loop biology and topological control in maintaining bacterial chromosome stability (Brochu et al., 2023). Ongoing investigations promise to uncover further fascinating insights into how perturbing this delicate balance precipitates genomic chaos. Ultimately, a deeper understanding of these processes in bacteria provides an invaluable foundation for elucidating common principles underpinning genome duplication, the most fundamental of biological transactions.

G-quadruplexes Formation

G-quadruplexes (G4s) are fascinating four-stranded DNA or RNA structures that readily self-assemble from guanine-rich sequences to form planar arrays of G-quartets stabilized by Hoogsteen base pairing (Maizels and Gray, 2013). G4 motifs enabling quadruplex formation are highly prevalent across genomes, occurring hundreds of thousands of times in human DNA (Linke et al., 2021; Miglietta et al., 2020). While *in vivo* functions of G4s were mysterious, compelling evidence now supports key functional roles for G4 structures. So how do G4s execute biological roles?

Intriguingly, over half the ~250,000 human DNA replication origins recently mapped correspond to G4 motifs. This implicates G4s in origin licensing, perhaps by recruiting initiator proteins. G4s are also emerging hotspots of genome instability when quadruplex-resolving helicases are defective (Figure 9). Common fragile sites often contain G4s, explaining their insta-

bility. Unrestrained G4 structures can thus pose hazards to genome maintenance (Maizels and Gray, 2013).

G4s clearly compose an influential regulatory and recombinogenic landscape controlling diverse genetic processes while enabling harmful instability if uncontrolled. Key future challenges are elucidating precise G4 structures that assemble *in vivo* and systematically delineating rules governing their functional protein interactions. Recent work has begun to elucidate the repertoire of helicases that resolve G-quadruplex structures in cells. Parekh and colleagues demonstrated key roles for RecG, DinG, and RecQ helicases in preventing G4-associated genetic instability in *E. coli* (Parekh et al., 2023). DinG deficiency increased instability of cloned G4 repeats, implicating DinG in removing transcription-driven R-loop associated quadruplexes. Meanwhile, RecQ mutation elevated instability while RecG mutation had the opposite effect, consistent with these enzymes resolving G4s during replication restart. These results provide direct *in vivo* evidence that bacterial helicases homologous to key human G4 processing enzymes help maintain genome integrity by resolving hazardous quadruplex structures. Defining the mechanisms of helicase targeting and processing of G4s in various genomic contexts remains an important goal toward understanding how cells harness these motifs while preventing deleterious outcomes.

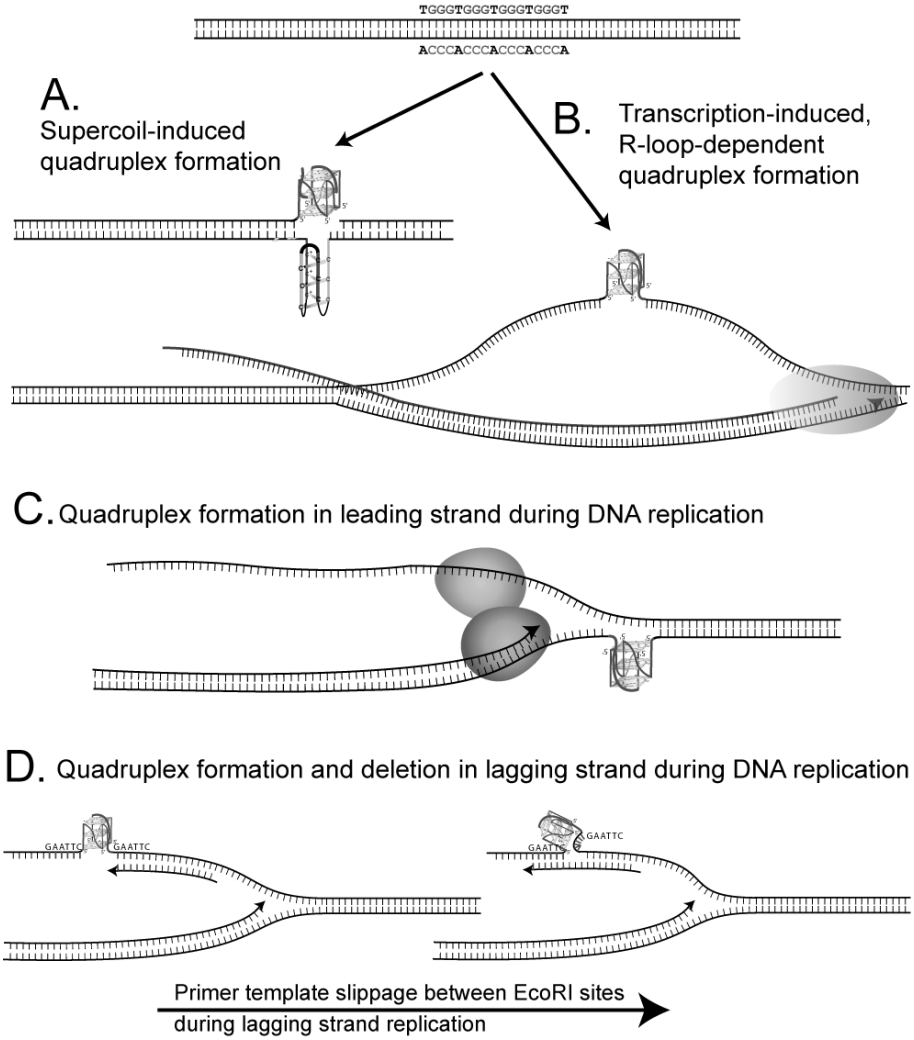

Figure 10: Alternative Pathways Resulting in Resolution of Quadruplex Structures. Quadruplexes may form spontaneously when DNA becomes supercoiled (Pathway A), in displaced strands subsequent to R-loop formation (Pathway B), in leading strands preceding replication forks if helicases unwind the strands far ahead of the fork (Pathway C), or in single-stranded lagging template strands during replication (Pathway D). In Pathway D, complete resolution likely occurs via replication slippage involving misalignment of primer and template between two proximal EcoRI sites produced when repeats are cloned into the *cat* gene. A stable G-quadruplex would be expected to obstruct replication (left), thereby promoting replication slippage (right). This is illustrated in the context of lagging strand replication, although it could occur during replication of either strand. (Reused from Parekh et al. (2023))

Termination

In organisms which replicate their chromosome bi-directionally, there will always be replisome-replisome collisions during the final stages of replication. This occurs only once for bacteria with one origin, such as *E. coli*, and 100s-1000s of time in eukaryotes with multiple linear chromosomes. Some bacteria possess a chromosome architecture that funnels replication into a particular region, termed the termination area, in which termination can be contained. Gram negative and positive bacteria have these systems and there is considerable variation in the layout and sequence specificity of these systems which essentially function for the same purpose. Not all bacteria possess these systems, such as *vibrio cholerae* which has two chromosomes and no known fork trap architecture (Galli et al., 2019).

Tus binds *ter* DNA in a polar manner through specific base interactions, the most notable interaction is the C6 base, locking into a binding pocket on Tus as a replication fork approaches from a non-permissible end. Early work by Brewer suggested that the role of the replication fork trap is to minimise replication-transcription conflicts (Brewer, 1988) and although that seems like a logical conclusion, it does not explain why other bacteria, such as *Vibrio cholerae*, lack a replication fork trap entirely yet still remaining a perfectly viable species (Galli et al., 2019). So why has this mechanism evolved in a subset of bacteria and what is its true function? The replication fork trap has been proposed to have been acquired through a plasmid integration into the chromosome, as Plasmid R1 contains an origin and two polar *ter* sites which coordinate replication of the plasmid, early studies highlighted that inactivation of the termination system through Tus deletion or *ter* sequence manipulation in R1 caused unstable plasmid maintenance and reduced genomic stability (Krabbe et al., 1997).

The first chromosomal fork block mechanism was discovered in *E. coli* through the integration of an ectopic replication origin into the chromosome (Kuempel et al., 1977; Louarn et al., 1979). By inactivating *oriC* using a temperature-sensitive version of DnaA and inserting an ectopic P2 prophage origin to drive replication initiation at a site 1 Mbp away from *oriC*, termination still occurred opposite *oriC* revealing that replication progression must be impeded in this area (Kuempel et al., 1977; Louarn et al., 1979). Sophisticated marker frequency analysis revealed specific termination or *ter* sites with the Louarn group identifying *terA* and *terC* and the Kuempel lab identifying *terA* and *terB* (Figure 11) (Hill et al., 1987; Massy et al., 1987). Developing *ter* consensus sequences allowed identification of additional sites, *terDEF*, using radioactively labelled probes (Hidaka et al., 1991). Once the *E. coli* genome sequence was available (Blattner et al.,

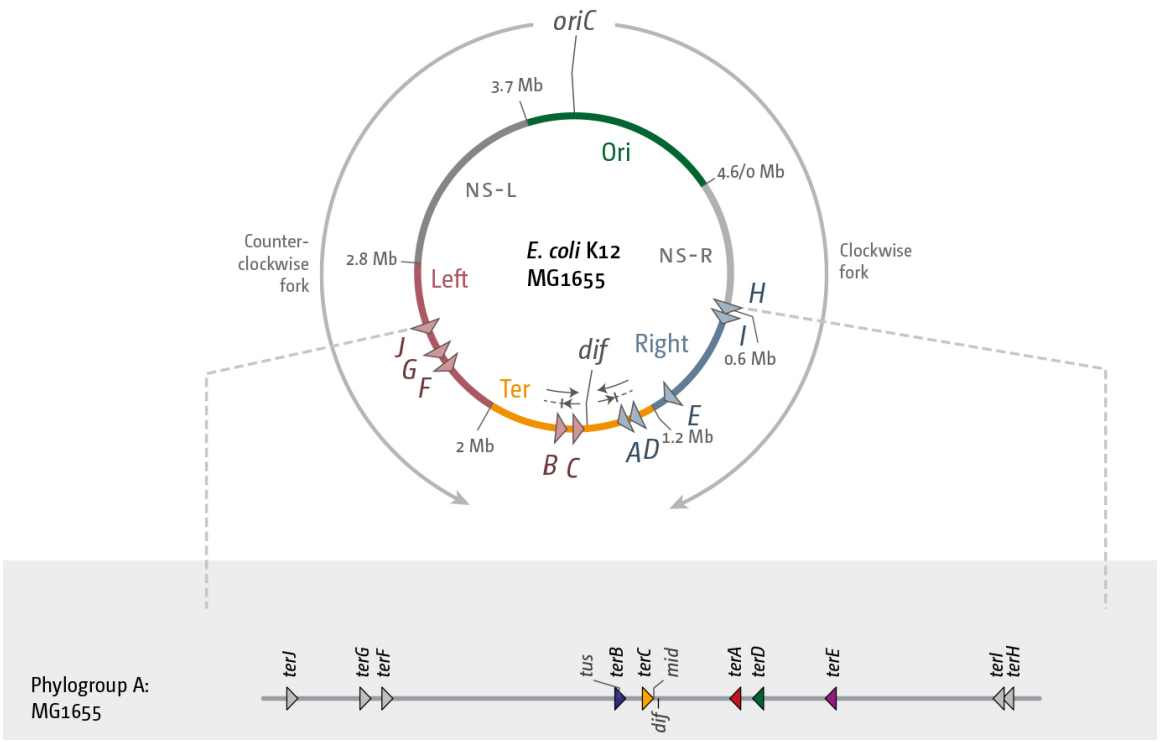

Figure 11: *E. coli* chromosome and fork trap architecture in MG1655. Polar *ter* sites are mapped as triangles, when replication forks encounter a tip of the triangle, they are stably arrested and replication can no longer proceed. The *dif* site is shown which is the final recombination site which decatenates chromosomal dimers. (Goodall et al., 2021)

1997), bioinformatics analysis identified *terG–terJ*. The *E. coli* chromosome is divided into two replichores with one replicated clockwise and the other counter-clockwise with five *ter* sites oriented to block clockwise replication fork complex while the other five block counter-clockwise replication fork complex (Figure 11) (Duggin and Bell, 2009). Further analysis has since shown that these later discovered *ter* sites are not to be considered part of the fork trap as they do not show active Tus binding or fork arrest activity (Duggin and Bell, 2009; Toft et al., 2021).

Two-dimensional (2D) agarose gel electrophoresis provides a powerful tool to directly visualize replication intermediates and fork fusion events at the molecular level. Duggin and Bell (2009) utilized this technique to analyze fork convergence and termination at *ter* sites in wild-type *E. coli* cells. They observed specific replication fork arrest and fork fusion intermediates accumulating at *terA*, *terC*, *terD*, and *terF* sites. In contrast, fork arrest or fusion events were not detected at the *dif* site, supporting conclusions from marker frequency analysis that *ter* sites mediate site-specific replication termination while *dif* does not. These 2D gel results provide direct physical evidence that converging replication forks fuse within the *ter* macrodomain to complete replication, validating models proposing the fork trap mechanism. This underscores the utility of 2D gels to elucidate mechanisms of fork fusion during termination at the DNA level.

The findings of specific fork arrest and fusion events at *ter* sites but not *dif* strengthen the model that *ter* sequences constitute a true replication fork trap in *E. coli*.

E. coli and *B. subtilis* are Gram negative and Gram positive bacteria, respectively, that both have replication fork trap systems that halt the movement of replisomes attempting to escape the termination area. Unlike *E. coli*, the terminator protein RTP in *B. subtilis* is a dimeric symmetrical protein that forms a winged-helical conformation when bound to genomic *ter* sites that arrests replication fork movement (Solar et al., 1998). The relative coverage of the chromosome occupied by *ter* sites, therefore defining the termination area, is also much more narrow in *B. subtilis* at roughly 5% of the chromosome, compared with the 20% coverage in *E. coli*. There have been extensive studies regarding the architecture of the RFT in Enterobacteriales that suggest that the diverse nature of the *E. coli* and *B. subtilis* RFT systems has likely been a result of convergent evolution as neither Tus or RTP share any significant sequence homology, possibly derived from plasmid origin and could help explain why these systems are seen in only 4 clades of bacteria (Galli et al., 2019).

Berghuis (Berghuis et al., 2015) has shown through the single molecule magnetic tweezer experiments on hairpin DNA, that Tus binds in a polar manner to *ter* DNA (they used *terB*) and the main mechanism of blocking the unwinding of DNA was the conserved C6 base which locks into a pocket and binds with crucial residues H144 and F140 (seen in Figure 12), supporting the initial findings of a molecular mouse trap (Mulcair et al., 2006). They clearly state that the protein-protein interaction between Tus and DnaB helicase is not needed to prevent unwinding of DNA and that the E49 residue shown in previous research (Mulugu et al., 2001) probably assists the formation of C6 binding through guiding the DNA as it starts to unwind, from the blocking orientation (Figure 12). E49K did reduce the probability of forming a successful locked complex, but that complex had the same strength as WT when it did form, implying a guidance. They finish by saying that if a helicase was present, this could perhaps increase the probability of C6 binding into the pocket, but that Tus-DnaB interaction is not needed in itself to cause blockage of an approaching replication fork.

The tweezer experiments show that an extreme amount of force is needed to overcome Tus-*ter* barriers in the blocking orientation, roughly 60pN, whereas unwinding and displacing Tus in the permissive direction required only 16pN. The *E. coli* replisome will unwind DNA at roughly 1kb/s (at its fastest) and Berghuis performed this experiment at roughly a rate of 30kb/s. This shows that the *E. coli* replisome is unlikely to overcome Tus in the blocking formation by force

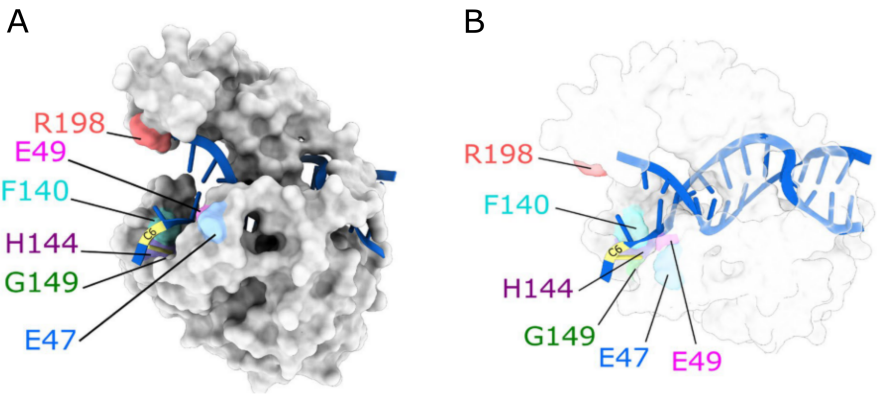

Figure 12: A) Opaque view of Tus-*ter* binding. B) transparent view of Tus-*ter* interactions to more clearly view the DNA unwinding and C6 binding pocket of Tus. Important residues are highlighted which are known to assist in maintaining the C6 binding pocket and orient ssDNA once DnaB helicase begins to unwind *ter* DNA from the non-permissible end. (Goodall et al. (2023))

alone and strengthens the idea that recombination will be needed to overcome the barrier and could explain why *ter* sites have been identified as recombination hotspots where repair mechanisms need to restart forks (Cox et al., 2000; Horiuchi et al., 1994; Hyrien, 2000; Kuzminov, 1995). Studies have shown how effective the Tus-*ter* complex is blocking approaching replication forks and permanently arresting them *in vitro*. However, studies using single molecule analysis and flow stretching assays which do not contain the replisome only show biochemical results which may only provide a piece to the puzzle of the events transpiring *in vivo* (Bastia et al., 2008; Berghuis et al., 2015; Mulugu et al., 2001).

The earlier data failed to consider factors such as the speed of the replisome in vivo and DNA supercoiling, which can affect DNA unwinding (Elshenawy et al., 2015). Elshenawy and colleagues used single-molecule DNA stretch flow assays to observe that Tus can be overcome by a fast-approaching replisome moving at speeds greater than 1000 bp/s, approximating *in vivo* speeds. This finding explains the *in vivo* data showing that only 50% of replication forks become arrested at Tus-ter sites (Elshenawy et al., 2015). Slow-moving forks, however, allow time for Tus to remodel ter DNA interactions, leading to the crucial C6 lock. When the C6 lock was preformed before a fast-approaching fork met Tus, forks were halted 89% of the time, whereas without the preformed lock, successful arrest was greatly reduced. The study also revealed the

role of residue R198, which interacts with the A5 residue of Tus; if this interaction is not permitted due to increased fork speeds, the chances of C6 locking are also reduced (Elshenawy et al., 2015). These findings extend the "mousetrap" model into a multistep model involving E49 guidance, R198 rearrangements, and C6 locking (Berghuis et al., 2018).

It has been suggested that the purpose of the RFT is to maintain replication-transcription directionality which minimise the potential conflicts. While this feature is clearly important in the *E. coli* chromosome, it seemed more likely that the RFT has evolved to reduce the effect of pathological replication but restricting the area in which forks fuse to the termination area. It has been shown that inactivating RecG helicase leads to over-replication inside the termination area (Dimude et al., 2015; Midgley-Smith et al., 2018; Rudolph et al., 2010a; Wendel et al., 2017). The current working hypothesis is that fusion of two replication forks leads to the accumulation of 3' flaps that can be targeted by PriA helicase to unwind the 5' end of this structure to facilitate the initiation of another replisome. Following DnaB loading onto the lagging strand and the replisome forming, there will be a double strand end left over which will be targeted by RecBCD to degrade the blunt end until it reaches a *chi* site, which are spread out over the chromosome, even within the termination area (Heller and Marians, 2006; Smith, 2012). Upon reaching the *chi* site there is a conformational change in the nuclease of the RecD enzyme that changes the nuclease activity to favour the 5' end and leaving the 3' end in tact. The resulting product is a 3' flap bound to by SSB, and later RecA in a filament complex with the ssDNA. Upon RecA binding, the 3' end can be introduced into an opened sister chromosome duplex to form a D-loop, yet another target for the PriA replication restart protein (Cockram et al., 2015). This newly established replisome will move in the opposite direction to the original fork, setting up a pathological replication cascade. It has been shown that the RFT introduces major issues to replication when ectopic origins are added and some strains result in lack of viability unless the fork trap is inactive ($oriC^{-}\ oriZ^{+}\ recG^{-}$).

The observed over-replication within the termination area has been used to study the effects of SDR on the ability to duplicate the chromosome when *oriC* is inactive (Rudolph et al., 2013). Cells that have a thermosensitive DnaA initiator protein, which is inactive at 42 degrees Celsius, can survive this inactivation in *recG* cells only if *tus* is inactive. The SDR can then extend past the termination area toward *oriC* where they can then fuse. Issues replicating in the opposite direction can occur from collisions, especially with the highly transcribed *rrn* operons, so destabilization of the ternary structure of RNAP is required to reduce the consequences of these collisions and allows the replisomes to displace RNAP more readily. Taken together it seems likely that a

main function of the replication fork trap is to restrict the area of fork fusion events to minimise over-replication that could reduce genome stability by increasing the number of recombination processes (Michel et al., 2018; Wendel et al., 2014).

Fork Trap System Categorization

The replication fork trap system present among the *Enterobacteriaceae*, vary somewhat across species, with either more narrow or spread out *ter* sites. As with all fork trap systems, the terminator protein *tus* lays within one of the *ter* sites, in *E. coli* this is *terB*, and autoregulates its own expression as when Tus-*ter* complexes are formed at *terB* there can be no nascent expression of *tus*. Recently, *in vitro* work by Jameson and colleagues (Jameson et al., 2021) provided strong evidence that forks stalled at Tus-*ter* are arrested until an opposing fork is encountered at the other end of Tus, where forks will fuse and terminate. When replication forks fuse at Tus-*ter* there is a consistent under-replicated area of the DNA between 15-24bp, which likely occurs when the replisomes are disassembled at each fork and Tus is displaced. This study discovered that none of the currently identified replication proteins or recombination proteins were able to mitigate this under-replicated section, implying that a currently unidentified enzyme/system is processing the under-replicated DNA after fusions occur at Tus-*ter* to conclude successful replication. Clearly, more is still to be uncovered about the exact function and molecular mechanisms which govern replication termination at Tus-*ter*, and fusion of two replisomes in general to bring about successful termination.

Fork trap systems are said to be categorized into two lineages, Type 1 (TI) and Type 2 (TII) fork trap systems (Toft et al., 2021). The TI subtype is said to represent an earlier more ancestral system, with the TII system having evolved more recently. Species with a TI fork trap possess only one *ter* site neighbouring the terminator protein gene, whereas TII fork traps, have two. With one *ter* site either side of the terminator protein gene, there could be some added benefit of the the more recent system compared to the ancestral system, explaining the rise of the TII fork trap in more recently evolved species, such as *E. coli* and *Shigella*. Figure 13 shows the theory for how a fork trap system would have been adopted into bacterial chromosomes from plasmid origin (Galli et al., 2019). However, comparisons between efficiency of these categories of fork trap cannot be made as there has been no evidence to show that one system is more efficient than that of the other, or what the local gene expression effect would occur when the autoregulated gene is sandwiched between two *ter* sites. More research into fork trap evolution will be valuable in understanding the final stages in DNA replication and we are only now beginning to understand the finer details of the evolution of fork trap system in bacteria (Goodall et al., 2023).

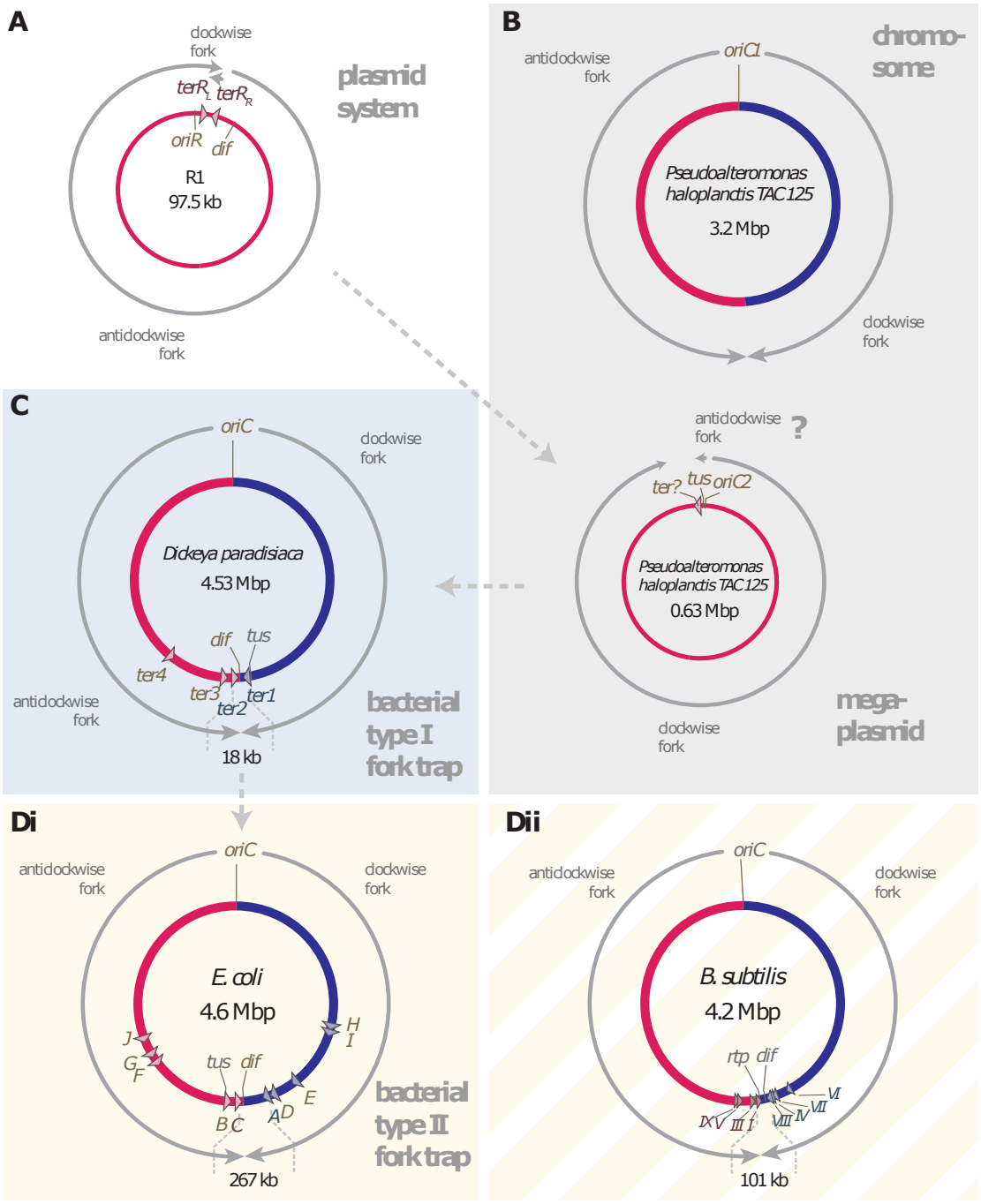

Figure 13: Illustration of fork trap systems in plasmids and bacterial chromosomes. (A) Replication dynamics and fork trap mechanism in the R1 plasmid. (B) Replication and fork trap features of the primary and secondary chromosomes in *Pseudoalteromonas haloplanctis* TAC 125. The secondary chromosome has a *tus* gene adjacent to *oriC2*. Existence of a *ter* site is speculative, if denoted by the question mark. (C) Type I fork trap system in *Dickeya paradisiaca*, described by Toft and colleagues (Toft et al., 2022). (D) Type II fork trap system in *E. coli* (Di) and the fork trap system in *B subtilis* (Dii). Refer to main text for additional details on fork trap evolution. (Taken from Goodall et al., 2023)

Accessory Helicases and Replication Termination in *E. coli*

Helicases play a vital role for all stages of DNA replication where their roles intersect with transcription and studies have highlighted importance of chromosomal structure in the coordination of the replisome (Zhang et al., 2023). Helicases in eukaryotes also mitigate end stage replication intermediates to prevent chromosome instability and fragility (Choudhary et al., 2023). Our deepened understanding of factors affecting DNA replication and repair will path the way for novel therapies and increase longevity for our own species, and lead to potential new antibiotics which can exploit unique aspects of the replication machinery to prevent disease progression from novel pathogens. Several inhibitors have been note which target specific domains on bacterial replicative polymerase PolC and DnaE to elicit antibiotic activity that lead to inability to fire new rounds of replication thus leading to cell death (Santos and Lamers, 2020).

Accessory helicases are important for processing these recombination intermediates (Brüning et al., 2014). They facilitate the replication of protein-bound DNA, which can often cause replication fork pausing or stalling. This is particularly important in the context of replication termination, where the replication fork must proceed smoothly to ensure accurate DNA replication and prevent genomic instability. In addition to facilitating replication termination, accessory helicases also play a key role in DNA repair. When replication forks encounter obstacles such as DNA lesions, the forks can stall and collapse. This requires a restart of the replication process, which is mediated by homologous recombination. Accessory helicases, such as RecG in *E. coli,* are crucial for processing the recombination intermediates and preventing pathological chromosomal amplification, thereby maintaining genomic stability.

Helicases such as RecG and UvrD are essential for both replication termination and DNA repair in *E. coli* (Epshtein et al., 2014; Lloyd and Rudolph, 2016). They ensure the smooth progression of the replication fork, facilitate the restart of stalled forks, and prevent chromosomal abnormalities, thereby maintaining the integrity of the *E. coli* genome.

RecG, for instance, is crucial for processing recombination intermediates and preventing pathological chromosomal amplification. This is particularly important when replication forks encounter obstacles such as DNA lesions, which can cause the forks to stall and collapse. In such cases, a restart of the replication process is required, which is mediated by homologous recombination (Rudolph et al., 2009). Similarly, UvrD plays a significant role in DNA replication and repair. It functions in nucleotide excision repair (NER) and helps remove RNA polymerase that has stalled at DNA damage sites, allowing repair synthesis to continue. UvrD can also disman-

tle RecA filaments formed during homologous recombination (Petrova et al., 2015; Veaute et al., 2005), a trait also shared by RecG helicase. Removing RecA filaments prevents unnecessary recombination during replication fork repair and maintains genomic stability. Therefore, studying accessory helicases involved in these processes is crucial to understanding mechanisms of DNA replication termination itself.

The role of RecG accessory helicase

RecG has been studied extensively for over 30 years and yet its precise function has not been fully realized. Initial ideas about how RecG can function in recombination and DNA damage repair came from early studies showing that cells lacking the Holliday junction resolvase RuvC are dependent on RecG for survival (Lloyd, 1991). RecG helicase plays a multifaceted role in promoting replication fork progression and stability in *E. coli*. As outlined in the introduction, RecG possesses branch migration and fork regression activities that allow it to remodel various DNA structures formed during replication, recombination and repair.

A key early finding was that RecG can regress model replication forks *in vitro* by simultaneously unwinding the nascent leading and lagging strands and reannealing the parental strands (McGlynn and Lloyd, 2001). This reversibility of fork structures likely underlies RecG's ability to generate Holliday junction substrates for RuvABC resolvase from stalled forks. The early work performed by Lloyd and colleagues also found RecG and PriA to be intimately linked in repairing DNA damage, through mutation analysis of PriA they identified allele changes which inactivated PriA and RecG showing their activity depends on helicase activity to process DNA intermediates (Al-Deib et al., 1996; Mahdi et al., 2003, 2012). This research formed the ground work for showing that RecG is a dsDNA translocase part of the SF2 helicases, which does not have any indication of possessing nuclease activity (Singleton et al., 2001).

in vivo, RecG helps maintain rapid and processive replication by facilitating fork movement through nucleoprotein barriers like transcribing RNA polymerases. Accessory helicases like RecG translocate along single-stranded DNA and displace proteins ahead of the fork (Rudolph et al., 2010b). RecG has specific roles in DNA repair facilitate the unwinding of D-loops and Holliday junction structures, seen in Figure 14 (Lloyd, 1991; McGlynn and Lloyd, 1999, 2000) and R-loops, which form behind RNA polymerases at sites of replication-transcription conflicts (Hong et al., 1995).

When replication forks do stall, perhaps at DNA lesions or difficult to replicate sequences, RecG

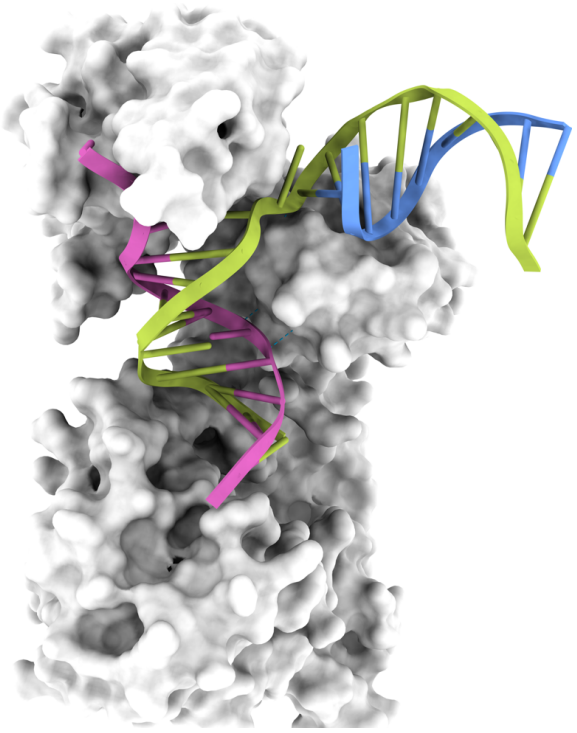

Figure 14: Model of RecG bound to replication fork, showing the parental and nascent strands, the critical 'wedge' domain of RecG can be seen separating the nascent DNA The protein-DNA complex was visualized in ChimeraX (PDB: 1GM5)

plays a key role in fork restart through its remodelling activities. Fork reversal and Holliday junction formation, either spontaneously or catalyzed by RecG, can facilitate lesion repair and replisome reloading (McGlynn and Lloyd, 2001; Rudolph et al., 2009). The regression activity of RecG likely promotes access to DNA damage and template switching mechanisms to bypass lesions (Rudolph et al., 2010a).

RecG has also been implicated in pathological events during defective replication termination. In cells lacking RecG, unrestrained over-replication occurs within the termination area between converging forks (Rudolph et al., 2013). This depends on PriA's ability to reinitiate replication at 3' flaps generated by improper fork fusion. RecG may normally convert these 3' structures into 5' flaps not recognized by PriA, preventing aberrant restart. Thus, RecG helps restrict pathological replication events spatially to the terminus region.

The work by Lloyd and colleagues showing that RecG is able to preferentially bind to 3' ssDNA structures and drive branch migration in the opposite direction to the replication fork (McGlynn and Lloyd, 1999). They showed that RecG can bind to three strand DNA structures such as D-loops and drive them back into a duplex DNA creating a Holliday junction which can then be resolved by RuvABC resolvase (Figure 15). Although here the authors state that RecG has no affinity for either 3' or 5' strand structures, later work clarified that RecG will preferentially

target the lagging strand of a stalled replication fork (Bianco, 2015; Briggs et al., 2003; Singleton et al., 2001). This fits well with more recent work suggesting that RecG has the ability to convert 3' flaps into 5' flaps within the termination area following fusion of replication forks (Midgley-Smith et al., 2018; Rudolph et al., 2013; Rudolph et al., 2019).

The preference of RecG for binding lagging strand template structures (McGlynn and Lloyd, 2001) suggests it may recognize specific fork geometries where leading strand synthesis has uncoupled and stalled ahead of the lagging strand. This could target its remodelling activities to forks requiring restart. The ability of RecG to unwind the leading strand of model forks despite translocating 3'-5' on single-stranded DNA (McGlynn and Lloyd, 2001) indicates it can move with opposite polarity on parental strand templates to coordinate unwinding of both fork arms.

RecG acts at several junctures to maintain replication integrity. It suppresses conflicts with transcription, processes blocked forks to template switch and restart replication, and prevents illegitimate restart events during termination. This multifunctionality underlies its role as a guardian of genome stability in *E. coli*. Ongoing work to improve the mechanistic understanding of RecG will provide further insight into the interplay between recombination and replication. Elucidating its diverse replication maintenance functions remains an important goal to fully define the complexities of bacterial chromosome duplication.

Recent work has provided important insights into RecG's mechanism of action and its crucial function in preventing aberrant DNA amplification events. Azeroglu and colleagues show that RecG prevents abnormal amplification flanking double-strand breaks (DSBs) undergoing repair and at stalled replication forks in *E. coli* (Azeroglu et al., 2016). They propose a model where RecG remodels branched DNA structures like D-loops, Holliday junctions, and stalled forks to enable correct PriA helicase loading. This directs proper replication restart instead of backward fork movement that causes over-replication.

Complementing these findings, Azeroglu and Leach reviewed models explaining how RecG suppresses genomic amplification. They highlight that while RecG can catalyze fork reversal biochemically, evidence *in vivo* is lacking. Instead, genetic analysis indicates RecG acts on branched structures like stalled forks to permit accurate PriA binding and helicase loading. This "reverse-restart" model posits that without RecG, PriA binds aberrantly at D-loops or arrested forks, driving backward replication and DNA over-amplification (Azeroglu and Leach, 2017). Together, these studies significantly advance understanding of RecG's critical activity in regulating replication restart. Remodeling branched DNA to direct proper PriA binding and replication fork

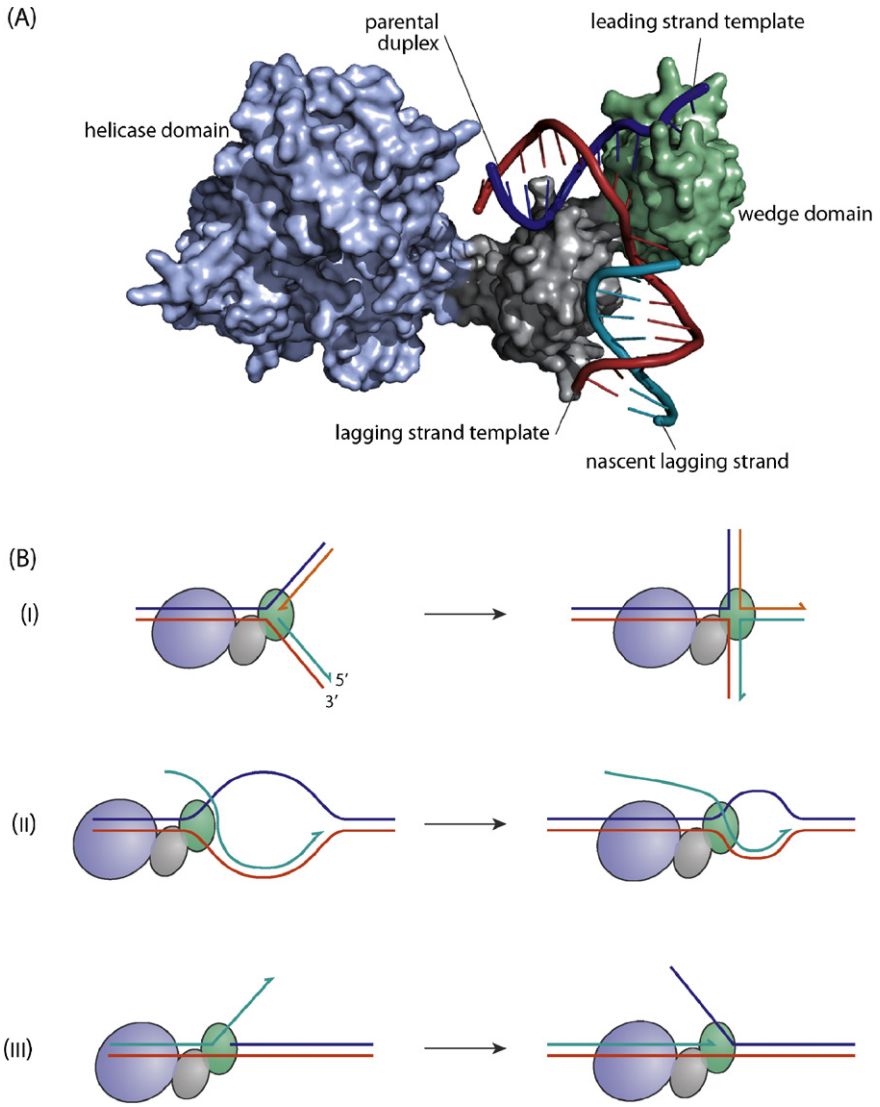

Figure 15: A) Molecular view of RecG and varying domain strcutures bound to replication fork. B)The pathways involving RecG have been shown to reduce recombination reactions by removing D-loop and R-loop structures on the one hand, yet form structures which act as recombination substrates, namely Holliday junctions. The function of RecG has been revealed by its 3D structure and what is termed the 'wedge' domain, able to prevent annealing of the nascent strands with the suggested mechanisms being to reverse stalled replication forks and reanneal the parental strands. Adapted from Rudolph et al, 2010)

movement emerges as RecG's key function at the nexus of replication and repair. Elucidating the molecular details of this Remodeling process remains an important goal. The conservation of RecG functions also raises exciting questions about control of replication dynamics in higher organisms.

Work by Lloyd and colleagues provides important new insights that may revise our understanding of RecG's primary cellular functions (Lloyd and Rudolph, 2016). Through analysis of recombination, replication, and chromosome defects in *recG* mutants, they reveal that many anomalies likely arise indirectly from failure to prevent pathological events like fork collisions and over-replication in the terminus region. Rather than directly catalyzing recombination or fork restart, RecG may mainly act to restrict excessive replication and recombination caused by fork problems (Figure 15). By quantitatively assessing DNA content and morphology, Lloyd et al. support a model where RecG suppresses DNA over-replication and excessive recombination events, preventing escalation into overt chromosome abnormalities. Their elegant epistasis analysis implies diverse *recG* mutant phenotypes result indirectly from loss of RecG's key function in controlling pathological DNA transactions at stalled and colliding replication forks. These novel insights compel re-evaluation of prior phenotypic data on RecG, emphasizing its central role in averting illegitimate DNA synthesis events by spatially confining pathological replication to fork trap zones. Overall, the literature presents an exciting new paradigm for understanding the pleiotropic defects of *recG* mutants in terms of RecG's critical purpose in preventing aberrant replication (Lloyd and Rudolph, 2016).

Outside of RecG's role in DNA repair, studies have unveiled RecG's unexpected multifunctionality as a regulator of gene expression, whereby its ability to remodel DNA structure enables RecG to directly control transcription at target promoters or assist binding of transcription factors like OxyR to induce gene activation (Yeom et al., 2012). This showcases RecG's diverse regulatory capacities and expands our understanding of how its helicase and branch migration activities influence gene expression networks (Heo and Park, 2015).

The role of UvrD accessory helicase

The *Escherichia coli* protein UvrD is a 3' to 5' DNA helicase that processes substrates containing both double-stranded and single-stranded DNA structures, translocating along and unwind DNA in its preferred 3' to 5' direction (Lee and Yang, 2006; Ordabayev et al., 2018), as seen in Figure 16. Early work showed that UvrD is essential for UV-induced DNA damage repair (Kuemmerle and Masker, 1980; Oeda et al., 1981; Todd and Glickman, 1979) and *uvrD* mutants induce hyper-recombination dependent on RecA filamentation (Arthur and Lloyd, 1980; Sargentini and Smith, 1981), which is mediated by the RecFOR pathway and upstream activity of *lexA* gene activity (Lloyd, 1983). Later work illustrated that UvrD functions as dimers and tetramers (Wollman J. et al., 2023; Yokota, 2020) and has a crucial role in removing RecA filaments that form on single-stranded DNA. It disassembles these filaments by translocating 5' to 3' along the DNA, countering the directionality of RecA polymerization and processes recombination intermediates like Holliday junctions, unwinding them to reverse strand exchange (Petrova et al., 2015). This helicase activity allows UvrD to act on stalled replication forks, where it clears RecA bound to lagging strand templates. By removing RecA, UvrD enables reversed fork structures to form, in which the fork unwinds itself into a 4-way junction. The reversed fork contains two double-stranded DNA arms and two single-stranded DNA arms that UvrD can further process (Wollman J. et al., 2023).

UvrD performs an essential fork-clearing function at stalled replication forks in DNA polymerase III mutants. Genetic analysis shows RecFORJQ proteins promote inappropriate RecA filament assembly on the lagging strand template at blocked forks (Florés et al., 2005; Veaute et al., 2005). Despite normally facilitating recombinational DNA repair, these RecA filaments are non-productive when polymerases are defective. The filaments cannot catalyze fork reversal, which requires template switching that RecFORJQ may inhibit. Instead, bound RecA persists and obstructs proper fork reversal and restart. UvrD, whose helicase activity displaces RecA from single-stranded DNA (Florés et al., 2005). By removing RecA, UvrD enables stalled forks to reverse independently through some alternate mechanism. This fork reversal is a critical early step in replication restart. Without UvrD to clear RecA, cells lose viability when replication enzymes are impaired and fork processing is disrupted. Thus UvrD serves a vital anti-recombinase role, not just to suppress hyper-recombination, but to recycle toxic RecA structures so replication can resume after fork blocks. While RecA filaments are crucial for these processes to take place, regulation of RecA filamentation and displacement is crucial to maintain genomic stability

(Florés et al., 2005).

The multifunctional helicases RecG and UvrD act as accessory factors to facilitate replication fork progression in bacteria. As described for RecG, these helicases possess diverse activities that help maintain replication integrity. Recent work has illuminated the importance of accessory helicases in resolving conflicts between replication and other processes like transcription and repair. Both RecG and UvrD can promote fork movement past nucleoprotein barriers by utilizing their translocase and unwinding activities, with UvrD being shown to displace Tus from *ter* DNA (Bidnenko et al., 2006; Hiasa and Marians, 1992). Additionally, UvrD plays key roles in processing stalled replication forks through barriers to enable completion of genome duplication. While RecG uses its capacity to promote Holliday junction resolution, UvrD partners with repair complexes to access lesions and recruit restart proteins (Atkinson et al., 2009; Wollman J. et al., 2023). Both RecG and UvrD show interactions with SSB and can dismantle RecA filamentation present on ssDNA to prevent unwanted recombination (Bianco, 2015; McGlynn and Lloyd, 1999; Rudolph et al., 2010b; Veaute et al., 2005). Though their mechanisms differ, RecG and UvrD serve complementary functions as accessory motor proteins that preserve replication fork integrity. With its multifaceted functions at the intersection of replication, recombination, and repair, UvrD exemplifies the complex interplay between helicase activities and nucleic acid transactions required to successfully duplicate the *E. coli* chromosome.

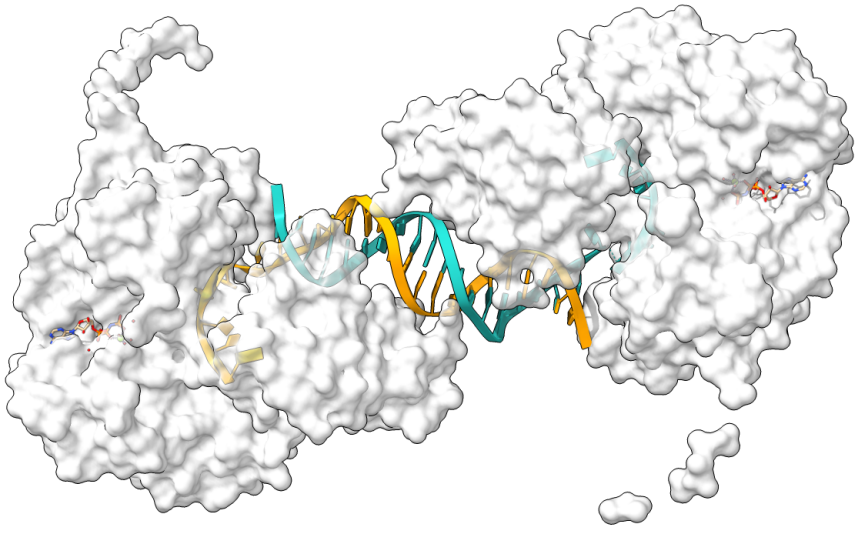

Figure 16: UvrD helicase utilizes a strand displacement mechanism to unwind double-stranded DNA in the 3'-5' direction powered by ATP binding and hydrolysis. Crystal structure depicts translocation along single-stranded DNA to displace the complementary strand, enabling UvrD to unwind DNA at sites of damage during nucleotide excision repair. Crystal structure from (PDB: 2IS4) (Lee and Yang, 2006)

One key function of UvrD is to promote fork movement through transcribing RNA polymerases. During transcription elongation, UvrD can bind to RNA polymerase and induce backtracking, forcing the polymerase to slide backwards along the DNA template (Epshtein et al., 2014). This helicase-driven backtracking exposes any DNA lesions shielded by stalled RNA polymerase, enabling nucleotide excision repair (NER) enzymes to access and repair the damaged sites. UvrD works together with the NER factor NusA to stimulate RNA polymerase backtracking and recruit repair proteins to exposed lesions. By scanning transcribing RNA polymerases and inducing backtracking, UvrD acts as a global damage surveillance factor that allows polymerases to detect lesions without terminating transcription. This mechanism for UvrD to clear nucleoprotein barriers ahead of replication forks is conserved across bacteria and eukaryotes (Epshtein et al., 2014) and showcases the importance of helicases being associated with the replisome in order to be directed to sites of DNA replication blocks or damage (Wollman J. et al., 2023).

In addition to backtracking stalled transcribing RNA polymerases, UvrD can also act directly on replication fork blocks like Tus-*ter* complexes (Hiasa and Marians, 1992). Ectopic Tus-*ter* sites integrated into the *E. coli* chromosome render both the SOS response and RecBCD recombination pathway essential for viability, as converging replication forks become arrested (Bidnenko et al., 2006). It has been shown that UvrD is also required for growth of SOS-deficient strains blocked by Tus-*ter*, likely by actively removing Tus from *ter* sites to liberate arrested replisomes (Bidnenko et al., 2006). Expression of UvrD or its *Bacillus subtilis* homolog PcrA is sufficient to restore viability, indicating a conserved antitermination function. At chromosomal Tus-*ter* sites, UvrD probably acts in concert with recombination proteins to process blocked forks after removing Tus .

In nucleotide excision repair initiated at strand discontinuities, UvrD partners with UvrAB to unwind DNA surrounding lesions (Atkinson et al., 2009). Although UvrAB has limited DNA unwinding capacity on its own, the UvrAB complex strongly stimulates UvrD-catalyzed unwinding of substrates containing nicks or gaps. UvrAB is proposed to increase UvrD recruitment or processivity at strand breaks. This stimulation helps explain the inviability of *uvrABD* mutants, as UvrD cannot efficiently unwind DNA during repair synthesis in the absence of UvrAB (Atkinson et al., 2009).

UvrD and Rep, although sharing homologies, antagonize the unwanted RecA activities that can occur stochastically, by removing RecA filaments formed on ssDNA, helping to preserve genomic integrity. Using its translocase activity, UvrD can actively displace RecA in an ATP hydrolysis-

dependent manner without needing RecA ATPase activity (Petrova et al., 2015). Efficient RecA displacement relies on accessible filament ends and availability of DNA binding sites for UvrD. UvrD likely removes RecA subunits from filament ends without fully translocating along the displaced strands. By dismantling RecA nucleoprotein filaments, UvrD helps prevent aberrant recombination events during replication (Petrova et al., 2015).

UvrD helicase uses diverse mechanisms to facilitate replication progression past obstacles like Tus-*ter* (Bidnenko et al., 2006) and transcribing RNA polymerases (Epshtein et al., 2014). UvrD processes intermediates during DNA repair and prevents excessive RecA activities. Despite functional overlap with the helicase Rep, UvrD exhibits specialized interactions with replication and repair factors that underlie its roles in preserving genome integrity. *uvrD* mutants have known genomic instability phenotypes, with Urrutia-Irazabel and colleagues showing increased R-loop accumulation, measured by the S9.6 antibody, and the authors highlight the lethality of a *uvrD rnhA* double mutant. Interestingly, *recG rnhA* double mutation is also lethal to cells (Urrutia-Irazabal et al., 2021).

Recent studies have revealed UvrD plays important accessory roles at replication forks in bacteria like *E. coli*, facilitated by its ability to form tetramers. As detailed recently by the Leake lab (Wollman J. et al., 2023), UvrD is present at most replication forks even during normal growth. This is likely due to its recruitment to help resolve frequent impediments like transcribing RNA polymerases and DNA lesions. In contrast to the Rep helicase, UvrD does not directly interact with replisome components, but rather is recruited as needed to help deal with replication stress. Importantly, UvrD helps maintain fork progression in a manner distinct from Rep.

These accessory activities build upon UvrD's previously known functions in nucleotide excision repair and mismatch repair pathways (Wollman J. et al., 2023). Indeed, disrupting UvrD's roles in these DNA repair processes reduced its localization to replication forks. This further confirms UvrD is recruited to forks by the blocks and lesions it helps resolve. Supporting this model, Liu and colleagues showed UvrD preferentially unwinds model fork structures with single-stranded gaps, consistent with it acting on impeded forks before they regress (Liu et al., 2019).

In eukaryotic cells, cell cycle controls largely separate transcription and DNA replication, however, recent research has uncovered that certain genes continue to be transcribed even during S phase when DNA synthesis occurs. How then do cells prevent clashes between the replication and transcription machineries on these actively expressed genes? The DNA helicases Pif1 and SETX have emerged as key players that function to limit genomic instability in eukaryotes. Pif1

helicases are multifunctional enzymes that promote replication fork progression through diverse barriers like G-quadruplex structures, R-loops, transcription complexes, and even tightly bound proteins including telomerase (Malone et al., 2022). By resolving these impediments, Pif1 ensures efficient DNA synthesis and preserves genome stability, alongside helpers such as conserved RNase HI and Rrm3 (Pohl and Zakian, 2019). At the same time, Pif1 also ensures successful replication termination of converging forks and clears DNA-protein barriers (Steinacher et al., 2012). Through these combined roles, Pif1 facilitates smooth replication fork movement, allowing complete and accurate duplication of the genome. Intriguingly, defects in these R-loop resolving helicases have been linked to breast cancer and neurological diseases respectively (Brambati et al., 2015; Hawkins et al., 2019). These connections highlight the importance of properly coordinating replication and transcription, and how disruptions to this delicate balance can lead to human disease.

Taken together, these recent studies paint a picture of UvrD as a multifunctional helicase that employs its tetrameric unwinding activity at DNA replication forks, DNA repair processes, and likely additional chromosomal sites. UvrD seems to act as a first responder that is summoned when its resolving functions are needed to maintain genome and cell integrity. Future work with UvrD mutants can further dissect its intricate and coordinated interactions that facilitate faithful DNA duplication and repair as well as its role overlapping pathways with other accessory helicases, such as RecG (Arthur and Lloyd, 1980; Lloyd, 1991).

Pathological Over-replication

Fork restart and termination are intricately linked processes required for successful chromosome replication. Defects in restart pathways lead to impaired termination, as efficient restart is needed to process termination intermediates like reversed forks and 3' flaps that accumulate when replisomes fuse. The replication fork trap plays a key role by spatially restricting pathological restart events to the termination area. Unrestrained restart throughout the chromosome via PriA-PriB risks instability. At Tus-*ter* sites, PriA, PriB and other restart proteins likely remodel termination structures to reload replisomes stalled for prolonged periods. Overall, restart pathways are crucial for handling termination intermediates and completing replication. Mutants defective in restart show termination defects (Dimude et al., 2016; Midgley-Smith et al., 2018; Rudolph et al., 2013).

Previous research has provided a comprehensive model of how replication terminates. It has

been shown that as replication forks fuse, the DnaB helicase of one replisome can displace the polymerase of the replisome on the leading strand. This displacement can lead to the formation of a 3' flap structure, which is a substrate for various proteins and recombinases (Figure 17). If PriA is allowed to load nascent replisomes, a double-strand break (DSB) can occur as the replisome progresses outward from the fusion site, potentially leading to a cascade of PriA-mediated over-initiation (Midgley-Smith et al., 2018; Midgley-Smith et al., 2019). Work in the Rudolph lab has shown that the hyper-replication phenotype observed in *recG* mutants is dependent on the activity of the PriA helicase and its ability to bind 3' flaps generated during defective termination events (Midgley-Smith et al., 2018; Midgley-Smith et al., 2019). Introduction of a *priA300* mutation, which abolishes PriA's capacity to unwind 3' structures, suppresses over-replication in *recG* cells (Rudolph et al., 2013). Therefore enzymes such as RecG and 3' exonucleases can regress or degrade the 3' flap structure, reducing the likelihood of this over-initiation reaction. In cells lacking these enzymes, over-replication specifically within the termination area has been observed, with cells being trapped by Tus-*ter* barriers. This model provides a detailed understanding of the complex process of replication termination.

There appears to be a pathological nature to the replication mechanism if it is over-active and restricted to the termination area and although there has been extensive research on pathological replication within the termination area, the exact cause and mechanism behind this pathological replication cascade remains elusive. There is consensus, however, that there is a distinct peak of replication within the termination area when RecG is inactive, as measured by Marker Frequency Analysis (MFA). While other hypotheses are that the pathological replication is generated from forks that are stalled at Tus-*ter* complexes (Sinha et al., 2020), and Wendel and co-workers take the view that over-replication occurs when nascent chromosomes move past each other before becoming degraded (Wendel et al., 2020), Rudolph and co-workers have argued against these hypotheses by finding that over-replication persists even in *tus* mutants where *oriC* firing is inhibited by growing cells at restrictive temperatures using the *dnaA* allele, that replication of the chromosome depends on the absence of Tus and not its presence. If over-replication was indeed initiating at Tus-*ter* complexes then we would expect to see less over-replication and an inviable phenotype in cells which lack Tus when *oriC* firing is inhibited.

Work by Rudolph and colleagues provides important new insights into the mechanisms that prevent pathological events during replication termination in bacteria like *E. coli* (Rudolph et al., 2013). Using deep sequencing and growth assays, they demonstrated that collisions between converging replication forks in the terminus region generate abnormal 3' flap structures

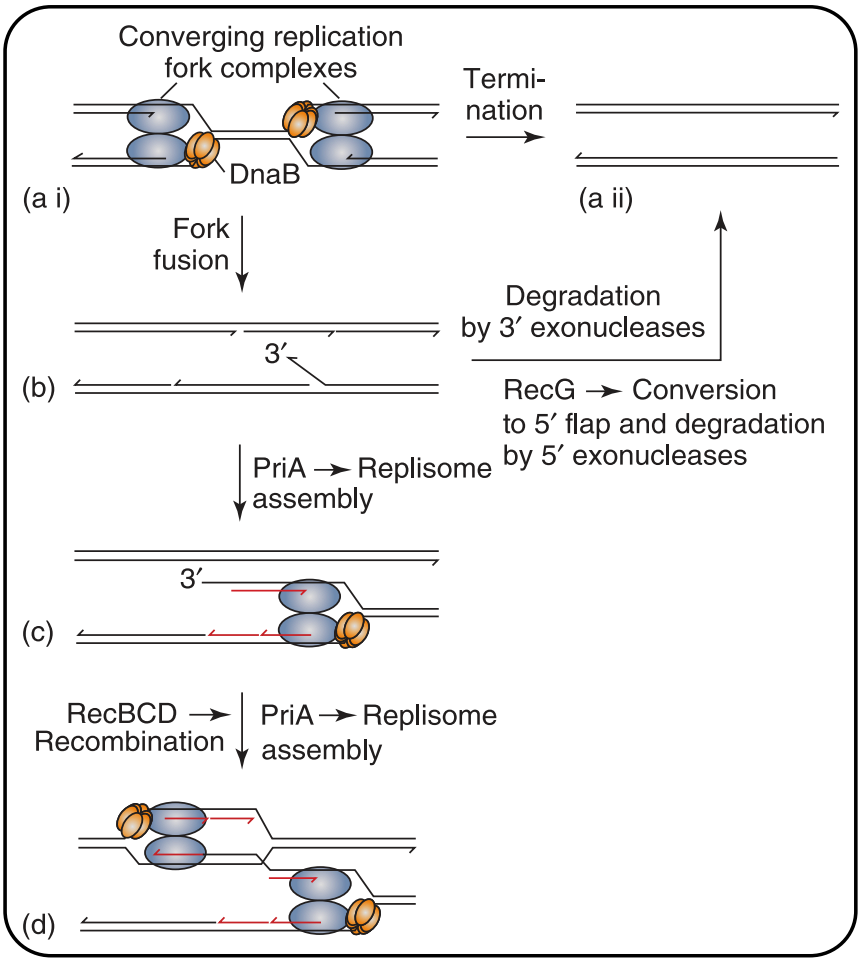

Figure 17: Diagram showing the hypothesis put forward initially by Lloyd and colleagues showing how replisome collisions can cause a 3' overhang which can be targeted by the restart protein PriA to initiate SDR within the termination area and begin a cascade of pathological replication. Repurposed from Rudolph et al, 2019)

that can be recognized by PriA helicase. This leads PriA to reload the replisome and drive undesirable re-replication of terminus DNA (Figure 17). Fork collisions also produce recombination substrates that elicit RecBCD-mediated formation of new replication forks. However, these pathological transactions only occur robustly in mutants lacking the accessory helicase RecG. In cells with intact RecG function, the helicase works together with exonucleases to constrain PriA and RecBCD activities on aberrant fork fusion structures. Linearizing the *E. coli* chromosome to prevent fork collisions also averts the pathological over-replication and recombination. These findings indicate that RecG plays a critical protective role during replication termination by restricting aberrant initiation and recombinational events on improper fork fusion intermediates. The conservation of factors like RecG, 3' ssDNA exonucleases and recombination pathways utilizing RecBCD and RuvABC, suggests overlapping control mechanisms are important to prevent genomic instability from fork collisions and altering the naturally processes at work during termination can lead to pathological over-replication (Dimude et al., 2016; Midgley-Smith et al., 2018; Rudolph et al., 2013).

Other work has documented that cells with a *recG* deletion will experience over-replication within their termination area, evident from Marker Frequency Analysis (MFA) showing the copy number against chromosome position in 1kb segments (Azeroglu et al., 2016; Wendel et al., 2014). Cells lacking RecG are also extremely sensitive to UV irradiation (Rudolph et al., 2009, 2010a) which incorporates lesions into the DNA, blocking replication fork progress. Over-replication can lead to head-to-tail collisions of replication forks, generating a DSE which can result in unwanted pathological recombination which increases chances of mutagenesis and introducing genomic instability, therefore it is is in cell's interest to minimized these events, unless the cell is trying to repair damaged DNA through recombination (Rudolph et al., 2010a). Two key hypotheses are discussed here that outline the possible mechanisms for this over-replication observed in cells lacking RecG and 3' exonucleases.

The Leach lab published data suggesting that forks within the termination area will be stalled at Tus-*ter*, awaiting the approach of the other fork and that, if the fork past a threshold, the replisome will disassemble and there will be a fork reversal and recombination in attempt to restart the fork. This goes wrong and leads to a cascade of replication that bounces around inside the termination area following a positive feedback loop of D-loop synthesis and Holliday junction accumulation, which will be lethal to the cell if they are not resolved (Azeroglu et al., 2016). They have shown that this form of initiation is dependent on RecA and RecBCD, therefore conclude that it is the repair pathway at Tus-*ter* which is responsible for the aberrant DNA synthesis in

the termination area. However, their model does not take into account that forks are not broken at Tus-*ter* complexes (Bidnenko et al., 2002) and will eventually need to fuse, and that this fusion may not always go as smoothly as once thought (Rudolph et al., 2013).

This model is in contrast with the Rudolph lab, who have supporting data indicating that the over-replication inside the RFT arises from the fusion of replication forks, which can lead to the accumulation of 3' OH flaps and in the absence of RecG and 3' exonucleases, will be the target of PriA to restart replication. These data show that wherever forks fuse, there is the potential for over-replication. Moreover, that cells can successfully replicate their genome from this form of initiation in the absence of *oriC*, but in contrast to the hypothesis put forward by Wendel and colleagues (Wendel et al., 2014), this depends on Tus being absent not present.

Work by the Rudolph lab have show that over-replication is mitigated when linearization of the chromosome occurs via *tos* integration and N15 bacteriophage lysogenic infection, abolishing any chance of fork fusions as replisomes will simply run off the ends of the linearized chromosome without fusing, essentially making the RFT redundant. Fluorescence microscopy analysis of replisome number also showed that there is a correlation between the inactivation of RecG and number of replisomes within the termination area in cells with a temperature sensitive DnaA protein and a deletion in the tus gene, forcing DNA duplication to originate within the termination area. These finding together support this model for over-replication seen in the termination area in Δ*recG* cells arises from fork fusions (Azeroglu et al., 2016; Rudolph et al., 2013). More defined techniques are needed to observe more detailed effects RecG has on bringing about successful termination.

The effect of RecG has been shown to be dependent on assembly of new replisomes by restart proteins PriA-PriB complexes within the termination area. Clear indication of this came from inducing mutant forms of PriA which either inactivate the helicase ability completely, or inactivate the ability to unwind specifically 3' flap structures (Rudolph et al., 2013). In these experiments it was shown that the over-replication present in recG cells was overcome with a *priA300* or *srgA1* mutations, as it was with the deletion of *priB*, the *priA300* allele encodes a protein which helicase activity has been abolished and *srgA1* encodes a protein deficient in unwinding structures resembling 3' flaps, strengthening the notion that when forks fuse there is the possibility of nascent fork synthesis at the 3' structure.

In all cases, the result would be an inhibition of nascent replisome assembly due to a lack of DnaB loading via the primosome. This supports the model that replication fork collisions can produce

3’ flaps which will be potentially targeted by PriA and initiate SDR within the termination area if RecG is not available to first convert the 3’ structure into a 5’ end, which is then degraded by 5’ exonucleases (RecJ), or 3’ exonucleases are able to degrade the 3’ structure before PriA or RecG are able to process the structure further (Midgley-Smith et al., 2018; Rudolph et al., 2013). Studies also observed that the *recG* phenotype is inhibited by *recB* mutations, evident from BrdU labelling experiments., showing recombination pathways may be at play to cause the over-replication. This is a logical conclusion seeing as the assembly of a nascent replisome from the fusion site of replication forks will essentially generate a double stranded end (DSE) and will be the site for RecBCD-RecA mediated HR forming a D-loop to allow further replisome assembly which will then move in the opposite direction to the original fork (Figure 17).

Despite advances in understanding RecG, its precise roles during replication termination remain incompletely defined. As an accessory helicase that acts on stalled forks, RecG may also function at fork fusion intermediates that arise during termination to facilitate proper merger and chromosome segregation. Elucidating these potential termination stage activities of RecG is an important goal for fully appreciating its multifaceted roles during DNA replication. Comparatively less is known about UvrD’s functions at terminating forks, presenting opportunities for new discovery into how its unwinding activities reshape fork structures to enable replication completion. Given their demonstrated importance as guardian helicases that counteract impediments throughout replication, further exploring the interplay between RecG and UvrD at termination will provide deeper insight into the mechanisms that preserve genome stability during the final stages of chromosome duplication.

In cells which have an additional ectopic origin, *oriZ*, present roughly half way in the right hand replicore, there is an increase in aberrant replication within the termination area when RecG is inactive, indicating that pathological replication depends on the amount of replication forks entering the terminus region. As Rudolph and colleagues suggest that it is the process of fork fusions that can drive the over-replication seen in *recG* cells which is lethal in cells with two origins. There is some speculation that cSDR could be a major player in the observed over-replication from R-loop formation at *oriK* sites, which are also centered around the termination area. This could make a compelling argument as RecG has been shown to degrade R-loops and in cells lacking RecG there has been shown to be accumulation of R-loops (Hong et al., 1995).

If R-loop persistence is driving the over-replication at *oriK* sites within the termination area when indeed, inactivating RecG would increase cSDR at this sites, work in the Rudolph lab has

shown that initiation is not taking place at R-loops in *recG* cells, whereas this likely is the case for *rnhA* mutants (Dimude et al., 2015). If over-replication is initiated at R-loops then this ectopic termination area would show a lack of over-replication in *recG* cells. However, a defined peak is present in this ectopic termination area strengthening the idea that the pathological over-replication cascade is R-loop independent and results from pathological events following fork fusion.

To investigate whether over-replication could be triggered at an ectopic termination area, Midgley-Smith and colleagues constructed an artificial fork trap in double-origin *E. coli* cells by integrating two *terA* sequences flanking the ectopic fusion point (Midgley-Smith et al., 2018). They found that while the ectopic fork trap had little effect on the replication profile in wildtype cells, it resulted in a defined peak of over-replication in *ΔrecG* mutants, in line with the over-replication seen in the native termination area in *recG* cells (Rudolph et al., 2010a, 2013). This supports the model that pathological replication is initiated by fork fusion events, rather than at specific 'hotspots' like *ter* sites (Horiuchi et al., 1994).

There is evidence that replisomes are stabilized at Tus-*ter* (Bidnenko et al., 2002; Moolman et al., 2016) and it is possible that this stabilization has evolved to reduce the likelihood of fork degradation that can lead to aberrant replisome assembly mediated by RecBCD. Indeed, if replisomes are stably halted at Tus-*ter* complexes then the DNA is essentially protected from degradation and unwinding by nucleases and helicases leading to recombination dependent fork restart (Azeroglu et al., 2016; Bidnenko et al., 2002; Dimude et al., 2018a; Rudolph et al., 2010a). However, replisomes are able to eventually overcome the Tus-*ter* barrier after being stalled. In double origin mutants the clockwise fork coming from *oriZ* would be the first to enter the native RFT area and become stalled at Tus-*ter* where the fork will need to wait for the opposing fork for some time before fusion can take place and it would appear that when forks have been stalled for a long period there is a chance of replisome disassembly which could trigger a 'naked fork' structure which could be processed by RecBCD leading to a 3' flap after fusion with the opposing fork. This 3' flap would then usually be targeted by RecG and exonucleases to prevent aberrant replication events. In the absence of RecG, however, there would be PriA mediated replisome assembly that is pathological.

Quadruple mutants carrying the *recG tus dnaA(ts)* rpo^* genotype are viable at restrictive temperatures showing that inactivation of the replication fork trap is essential for survival in cells that lack origin firing capabilities and that transcription-replication conflicts need to be miti-

gated in order to allow replication to duplicate the entire chromosome when moving out from the terminus region toward *oriC*. In these cells, lethality has been reported when the RuvABC Holliday junction resolvase complex or RecA becomes inactive, showing a tight interplay between over-replication in the termination area an recombination. Without the HJ resolvase or RecA to mediate the strand invasion and formation of a D-loop, these recombination intermediates would accumulate in the termination area leading to poor chromosome integrity and ultimately cell death. This is further supported by reduced replication profile within the termination area in *priA300* mutants and the increased viability when Tus is inactive, which would presumably allow forks to proceed out of the RFT area and towards *oriC*, essentially spreading out the concentrated Holliday junctions across the chromosome that increases stability and may help bring recombination to equilibrium. Of course, Singleton (Singleton et al., 2001) showed that RecG can process HJs back into replication forks so there may be a compounding effect by inactivating RecG that allows accumulation of 3'OH flaps used for replisome assembly and also allows HJs to persist for longer periods.

The interaction between recombination and replication has now been well established, where multiple researchers point to recombination systems being crucial for the repair of DNA damage and the ability of cells to successfully restart the replication machinery after becoming blocked at the damaged sites or lesions (Bidnenko et al., 2002; Briggs et al., 2003; Kogoma, 1997; Kreuzer, 2005). While recombination proteins, RecBCD, are crucial for the final stages of replication (Sinha et al., 2017, 2018, 2020), there is good evidence pointing toward recombination being involved in the pathological replication observed in the termination region in *recG* cells, being dependent on RecA (Dimude et al., 2015; Rudolph et al., 2010c, 2010b, 2013). yet more work needs to be done to uncover the nuances of replication termination and involvement of accessory helicases such as RecG and UvrD, as well as other key players.

Aims and Objectives

My research aims for this PhD project are to further elucidate mechanisms governing proper DNA replication termination in *E. coli*. One goal is to further characterize the replication fork trap system that spatially restricts termination events and examine its conservation across Enterobacteriaceae species. I aim to further elucidate the consequences of the fork fusion event during termination, including effects on recombination rates at defined chromosomal sites. Additionally, I will investigate the mechanisms of lethality of *rnhA recG* double mutants by quantifying global R-loop accumulation in cells lacking these key enzymes, and any interplay between RecG and UvrD helicase. By improving understanding of termination control mechanisms by looking into the phylogenetic conservation of this system, its impact on recombinational rates and RNA:DNA hybrid accumulation, my work will provide novel insights into how replication completion is coordinated, and how the fork fusion process is inherently consequential that requires careful maintenance to regulate genomic stability. My motivating hypothesis is that specialized systems like the fork trap prevent pathological events during the final stages of replication, and loss of these mechanisms leads to unrestrained over-replication and DNA intermediate accumulation which increases genomic instability, and that accessory helicases such as RecG and UvrD are required for processing intermediates generated from fusion events. Elucidating these termination safeguard mechanisms in *E. coli* will advance knowledge of a critical yet still enigmatic phase of bacterial DNA replication.

The findings presented in this thesis advance our understanding of mechanisms at play during replication termination and highlight the genomic instability caused by fork fusions during the termination event. I cover the architecture of the *E. coli* chromosome and how the evolution of the replication fork trap provides a benefit to cells once the system has been acquired. I present novel chromosome location specific recombination rate data which highlights the inherent consequences of fork fusions during termination, and I show that R-loops are increased in mutants lacking key players that are needed to resolve the intermediates caused by fork fusions. I build on from the current model of intermediate 3' flap formation and propose an extension to the model which incorporates a RecA-independent mechanism of homologous recombination, all showing that even with the recent advances in this field, we still do not have a full picture of the molecular processes needed to complete DNA replication.

/newpage

Materials and Methods

Strains and Plasmids

Table 1: Strains

strain	Genotype	Source
DG001	*MG1655*	NA
DG002	*oriZ::cat*	RCe504 copy
DG003	*tus::cat*	RCe223 copy
DG004	*tus::cat*	JD1557 copy
DG005	*DH5a*	NA
DG006	*dam- N15 lysogen*	JD1593 copy
DG007	*PriA300*	RCe150 copy
DG008	*TB28 PriA300 RecG*	AU1014 copy
DG009	*narU-kankanMX4*	SLM1042 copy
DG010	*yjhR-kankanMX4*	SLM1043 copy
DG011	*narU-kankanMX4 tus::dhfr*	DG009 x p1RCe224
DG012	*yjhR-kankanMX4 tus::dhfr*	DG010 x p1.RCe224
DG013	*oriZ:FRT*	SLM1051 copy
DG014	*oriX:FRT TB28*	JD1338 copy
DG015	*OriZ:FRT TB28*	JD1339 copy
DG016	*W3110*	NA
DG021	*rnhA::dhfr*	RCe590 copy
DG022	*rnhA::apra*	RCe591 copy
DG024	*oriX:FRT TB28 yjhR-kankanMX4*	DG014 x p1.DG010

DG026	*oriZ:FRT TB28 yjhR-kankanMX4*	DG015 x p1.DG010
DG027	*oriX:FRT TB28 narU-kankanMX4*	DG014 x p1.DG009
DG028	*oriZ:FRT TB28 narU-kankanMX4*	DG015 x p1.DG009
DG029	*oriX:FRT TB28 yjhR-kankanMX4 tus::dhfr*	DG024 x p1. DG011
DG030	*oriZ:FRT TB28 yjhR-kankanMX4 tus::dhfr*	DG026 x p1.DG011
DG031	*lacLYZA- lacO34 rpo* uvrD::dhfr pAM407*	Rce233 copy
DG032	*OriX:FRT narU-kankanMX4 tus::dhfr*	DG027 x p1.DG011
DG033	*OriZ:FRT narU-kankanMX4 tus::dhfr*	DG028 x p1.DG011
DG034	*rnhA::cat recG::apra pDM104-recG*	JD1450 copy
DG035	*OriZ:FRT recG::apra*	DG015 x p1.DG008
DG036	*uvrD::dhfr*	WT x p1.DG031
DG037	*OriZ:FRT pAM407-uvrD*	DG015 x pAM407
DG038	*tus::cat recG::apra*	DG003 x p1.DG008
DG039	*OriZ:FRT uvrD::dhfr pAM407*	DG037 x p1.DG031
DG040	*recG::apra*	WT x p1.DG008
DG041	*OriZ::cat tus::dhfr*	DG002 x p1.DG011
DG042	*OriZ::cat recG::apra*	DG002 x p1.DG008
DG044	*OriZ::cat tus::dhfr recG::apra*	DG041 x p1.DG008
DG045	*tus::dhfr*	WT x p1.DG011
DG046	*Mcherry-cat*	RCe829 copy

Table 2: Strain continued

strain	Genotype	Source
DG047	*recG::apra tus::dhfr*	DG040 x p1.DG011
DG048	*uvrD::dhfr recG::apra*	DG036 x p1.DG008
DG049	*xonA::apra*	WT x p1.RCe563
DG050	*xseA::dhfr*	WT x p1.SLM1225
DG051	*Mcherry-cat*	WT x p1.DG046
DG052	*uvrD::dhfr recG::apra tus::cat*	DG048 x p1.DG003
DG053	*rnhA::dhfr pDIM104-recG*	DG021 x pMD104

DG054	*recG::apra pDIM104-recG*	DG040 x pMD104
DG055	*DH5? pSLM001-kankanMX4*	SLM1021 copy
DG056	*pKD46-ts*	WT x pKD46
DG059	*tldD-kankanMX4*	DG056 x PCR DG009
DG060	*tldD-kankanMX4*	WT x p1.DG059
DG061	*OriX:FRT tldD-kankanMX4*	DG014 x p1.DG059
DG062	*OriZ:FRT tldD-kankanMX4*	DG015 x p1.DG059
DG063	*recA::apra*	RCe446 copy
DG064	*narU-kankanMX4 recG::apra*	DG009 x p1.DG040
DG065	*narU-kankanMX4 recG::apra tus::dhfr*	DG011 x p1.DG040
DG066	*OriZ:FRT narU-kankanMX4 recG::apra*	DG028 x p1.DG040
DG067	*OriZ:FRT narU-kankanMX4 recG::apra tus::dhfr*	DG033 x p1.DG040
DG068	*yjhR-kankanMX4 recG::apra*	DG010 x p1.DG040
DG069	*OriZ:FRT yjhR-kankanMX4 recG::apra*	DG026 x p1.DG040
DG070	*tldD-kankanMX4 recG::apra*	DG060 x p1.DG040
DG071	*OriZ:FRT tldD-kankanMX4 recG::apra*	DG062 x p1.DG040
DG074	*rnhA::cat tus::dhfr*	Rce633 copy
DG075	*recA::apra*	WT x p1.DG063
DG076	*yjhR-kankanMX4 recA::apra*	DG010 x p1.DG063
DG077	*OriZ:FRT yjhR-kankanMX4 recA::apra*	DG026 x p1.DG063
DG078	*uvrD::dhfr narU-kankanMX4*	DG036 x p1.DG009
DG079	*uvrD::dhfr recG::apra narU-kankanMX4*	DG048 x p1.DG009
DG080	*xonA::apra narU-kankanMX4*	DG049 x p1.DG009
DG081	*xseA::dhfr narU-kankanMX4*	DG050 x p1.DG009

Table 3: Plasmids

Number	Description
pDIM104	*recG* expression via pAra promoter. pBAD24 background
pKD46	Recombineering plasmid, expression of Lambda phage genes exo, bet and gam
pCP20	FLP recombinase expression
pRS316-kankanMX4	Original kankanMX4 plasmid with pRS316 backbone
pSLM001	Clean kankanMX4 plasmid

Broth and Media

LB medium

Bacterial growth medium was prepared using Luria-Bertani (LB) broth consisting of 1% bacto tryptone (BD Biosciences), 0.5% yeast extract (BD Biosciences), 0.05% sodium chloride (Fisher), and 0.002 M sodium hydroxide (Fisher) at a pH of approximately 7.0. For solid medium, LB broth was aliquoted into 200 ml and 300 ml volumes, and agar (Sigma Aldrich) was added to reach final concentrations of 1.5% (3 g and 4.5 g respectively). The completed LB agar was then autoclaved and poured into Petri dishes under sterile conditions.

Mu medium

Bacterial growth medium (broth) was prepared containing 1% bacto tryptone (BD Biosciences), 0.5% yeast extract (BD Biosciences), 1% sodium chloride (Fisher Scientific), and 0.002 M sodium hydroxide (Fisher) at a pH of approximately 7.0. For solid medium, the broth was aliquoted into volumes of 200 ml and 300 ml, to which 1% agar (Sigma Aldrich) was added at 2 g and 3 g, respectively, prior to autoclaving.

M9 Minimal salt medium

M9 minimal media was prepared using DifcoTM M9 Minimal Salts base (BD Biosciences). The powder was used to make a 5X M9 minimal salts solution, which was aliquoted into 50 ml portions and autoclaved. The 5X solution was then diluted 5-fold, re-aliquoted into 50 ml portions, and autoclaved again to make a 1X solution.

For M9 agar, 50 ml of the 5X M9 minimal salts solution was added to 200 ml of molten 1.8% agar (3.6 g/L). The mixture was supplemented aseptically with final concentrations of 2 mM magnesium sulfate ($MgSO_4$), 0.1 mM calcium chloride ($CaCl_2$) (both from Sigma-Aldrich), 0.4% glucose, and 0.05% casamino acids once the temperature reached 50°C to prevent premature solidification.

The 1X M9 minimal salts solution was used for serial dilutions of bacterial and bacteriophage P1 lysate cultures. Immediately prior to use, the 1X solution was supplemented aseptically with 2 mM magnesium sulfate and 0.1 mM calcium chloride.

Salt differences and antibiotics

When performing various genetic techniques in *E. coli*, we have found that the salt concentration of the growth media impacts the effectiveness of different antibiotics. Specifically, we observe better results using Mueller Hinton (Mu) media, which contains higher salt levels, in combination with kanamycin or chloramphenicol for select experiments. Mu media is optimal when performing recombineering with chloramphenicol as the selective marker. Additionally, Mu media paired with kanamycin yields improved outcomes compared to low salt media when assaying recombination rates by selecting kanamycin-resistant revertants. In contrast, low salt media is preferable for transductions with apramycin selection, as the high salt Mu media allows breakthrough growth of apramycin sensitive cells, evidenced by heterogenous colony sizes. Taken together, our observations indicate that high salt Mu media is optimal for recombineering and selecting kanamycin or chloramphenicol resistant transformants or revertants. However, low salt media should be used for apramycin selection in transductions to prevent uncontrolled growth of untransformed cells. This was observed consistently throughout experiments and shows the importance of salt concentration in optimizing protocols for genetic techniques in *E. coli*. We also noticed that *oriX* strains grow better in low salt media, such as LB, compared to Mu, and when paired with accessory helicase deletions, in *recG* or *uvrD* for example, the strain requires extra day or two of incubation for healthy colonies to develop illustrating the nuance of altering salt concentrations and experimental outcomes.

Antibiotics and supplements

Table 4: Antibiotics and Supplements Table

Supplement	stock conc	final conc
Ampicillin	5 mg/ml	50 ug/ml
Apramycin	2 mg/ml	20 ug/ml
Chloramphenicol	1 mg/ml	10 ug/ml
Kanamycin	4 mg/ml	40 ug/ml
Tetracycline	10 mg/ml	10 ug/ml
Trimethoprim	50 mg/ml	10 ug/ml
Arabinose	20%	0.20%
Glucose	20%	0.20%

Buffers

MC Buffer

A 500 mM stock solution of calcium chloride ($CaCl_2$) was prepared by dissolving 5.549 g of $CaCl_2$ powder (Sigma-Aldrich) in 100 ml of deionized water. The CaCl2 was allowed to fully dissolve under gentle agitation.

For the magnesium sulfate ($MgSO_4$) solution, 0.6 g of $MgSO_4$ powder (Sigma-Aldrich) was measured using an analytical balance and transferred to a 50 ml graduated cylinder. Approximately 40 ml of deionized water was added to the cylinder and the MgSO4 was dissolved by capping the cylinder with Parafilm and gently inverting 5 times. Once fully dissolved, 500 µl of the 500 mM $CaCl_2$ stock solution was added to the MgSO4 solution. The total volume was brought up to 50 ml with additional deionized water. The contents were mixed by inverting the capped cylinder 5 times. The complete $MgSO_4$/$CaCl_2$ solution was sterile filtered through a 0.22 µm syringe filter into 5 ml aliquots and stored at 4°C until use. The final concentrations were 100 mM $MgSO_4$and 5 mM $CaCl_2$ based on the amounts and volumes used.

TBE Buffer

Tris-borate-EDTA (TBE) electrophoresis buffer was prepared as follows. A 10X concentrated stock solution was made by dissolving 108 g Tris base (Sigma-Aldrich) and 55 g boric acid in approximately 900 ml of deionized water to give final concentrations of 0.89 M Tris and 0.89 M boric acid. Next, 40 ml of 0.5 M EDTA (pH 8.0) was added to achieve a final EDTA concentration of 0.02 M. The solution was then adjusted to a final volume of 1 L to produce the 10X TBE stock solution. For experiments, a 1X working solution of TBE was prepared by diluting 100 ml of the 10X stock with 900 ml of deionized water.

S9.6 Antibody

The S9.6 monoclonal antibody, originally developed by Boguslawski et al. (Boguslawski et al., 1986), has been widely used to detect and map R-loop structures. S9.6 was generated by immunizing mice with a synthetic RNA:DNA hybrid and was shown to exhibit high affinity and specificity for RNA:DNA hybrids (Boguslawski et al., 1986). Subsequent studies demonstrated that S9.6 could immunoprecipitate R-loop structures, enabling their genome-wide mapping through techniques like DNA:RNA immunoprecipitation followed by sequencing (DRIP-seq) (Sanz et al., 2021).

While S9.6 has been instrumental in mapping R-loop structures genome-wide, recent studies have clarified its binding specificities and limitations. König et al. found that S9.6 exhibits variable binding affinities towards different RNA:DNA hybrid sequences, ranging from low nanomolar to micromolar Kd values (König et al., 2017). Furthermore, S9.6 shows cross-reactivity with RNA duplexes, which can confound certain applications. Smolka et al. revealed that the cytoplasmic and nucleolar staining patterns commonly observed with S9.6 immunofluorescence microscopy actually arise predominantly from ribosomal RNA rather than R-loops, and that pull downs with the S9.6 are RNase HI sensitive and RNase TI insensitive, providing further evidence that the structures in which the antibody bind are indeed R-loops (Smolka et al., 2021).

Despite these limitations, S9.6 remains a valuable tool for investigating R-loop biology when used with appropriate controls. Treating samples with RNase H, which specifically degrades the RNA strand of RNA:DNA hybrids, provides a key control to verify the hybrid-dependence of S9.6 signals (Sanz et al., 2021; Smolka et al., 2021). Furthermore, mapping methods that sequence the DNA component of immunoprecipitated R-loops, such as DRIP-seq and the higher-resolution sDRIP-seq approach, largely circumvent issues with RNA contamination (Sanz et al., 2021).

In our research, we have successfully employed S9.6-based methods to stain global R-loop structures following best practices described in the literature. By using these optimized approaches and necessary controls, we are able to leverage the power of S9.6 to detect R-loops in gDNA extracts while mitigating the potential limitations of the antibody by incorporating the necessary RNase controls. Performing DRIP-seq on these extracts was beyond the scope of this project.

Strain Construction

P1 Transduction

Genetic manipulation of the *E. coli* genome was predominantly performed as described by Thomason (Thomason et al., 2007). Briefly, P1vir phage lysates were prepared by infecting the donor *E. coli* strain containing the genetic marker of interest with P1vir phage. The donor strain was grown overnight in LB broth at 30-37°C. The next day, the culture was diluted 1:100 into fresh LB broth containing 0.2% glucose and 5 mM CaCl2 and grown to mid-log phase (A_{600} 0.3-0.6). P1vir phage was added at an MOI of 0.1-1 and the culture incubated until complete lysis (2-3 hrs). Cell debris was removed by centrifugation and the P1 lysate stored at 4°C with chloroform.

For transduction, recipient *E. coli* cells were grown overnight in LB broth, pelleted and resuspended in P1 salts solution (10 mM MgSO4, 10 mM Tris pH 7.4). Varying amounts of P1 donor lysate (1-100 µl) were mixed with 100 µl of recipient cells and incubated for 30 min at 37°C to allow phage adsorption. The mixture was then incubated in LB broth with 1M sodium citrate for 1 hr at 37°C before plating on selective LB agar plates containing 5 mM sodium citrate. Transductant colonies were purified by streaking on selective plates. Insertion of the genetic marker was confirmed by PCR using primers flanking the insertion site.

The distance between genetic markers can be estimated by the frequency of co-transduction. Markers less than 1 minute apart on the *E. coli* chromosome transduce together at frequencies above 90%. More distant markers are co-transduced at lower frequencies proportional to the distance between them.

Bacterial transformation

E. coli dh5α cells were grown in LB medium to an optical density (A_{600}) of 0.4. Cells from 11 mL of culture were harvested by centrifugation at 4°C and washed 5 times with ice-cold 10% glycerol. The cell pellet was resuspended in 1 mL of ice-cold 10% glycerol. 1 µL of the desired plasmid was added to the cell suspension and incubated on ice for 30 minutes. 50 µL of the cell/plasmid mix was transferred to a pre-chilled electroporation cuvette and electroporated at X volts for X milliseconds using an ECM630 electroporator (BTX Harvard Apparatus). Immediately after pulsing, 1 mL of SOC recovery medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM glucose) was added. The cells were

transferred to a 15 mL culture tube and incubated at 37°C for 1 hour with shaking at 225 rpm.

After the recovery period, 500 µL of the transformation mix was spread onto LB agar plates containing 50 µg/mL ampicillin and incubated overnight at 37°C. The remaining 500 µL was kept at room temperature overnight as a backup and plated the next day if no transformants were obtained on the initial plates. Transformants containing the plasmid were selected by ampicillin resistance.

Genomic Recombineering by Inactivating genes using PCR products

This method allows rapid one-step gene inactivation using short homology PCR products in *E. coli* strains expressing the Red recombinase and was performed essentially as described (Datsenko and Wanner, 2000). Linear PCR products containing antibiotic resistance cassettes flanked by homology regions to a target gene were generated using primers with 36-50 bp extensions homologous to the flanking regions of the gene of interest.

Recombineering of the *tldD* gene in *E. coli* was performed using this method. The strain DG056 containing the pKD46 plasmid with the lambda Red recombinase system was used as the host. An overnight culture of DG056 was grown in LB with ampicillin at 30°C to maintain the temperature-sensitive pKD46 plasmid. The next day, 100 µL of the overnight culture was inoculated into 11 mL of LB with ampicillin and 0.2% L-arabinose and grown at 30°C to an A_{600} of approximately 0.6. Expression of the lambda Red recombinase genes on pKD46 was induced with L-arabinose. The 11 mL culture was centrifuged at 5000 rpm for 5 minutes at 4°C to pellet the cells, which were then washed four times in decreasing volumes of ice-cold 10% glycerol (from 11 mL to 500 µL) to make the cells electrocompetent.

The cells were kept on ice for 30 minutes with 1 µL of purified PCR product containing the kanamycin resistance cassette flanked by 50 bp homology arms to the *tldD* gene. Electroporation was performed by mixing 50 µL of cells with the PCR product in a 0.1 cm cuvette and pulsing at 1.7 kV for 5.5 ms using an electroporator. The cells were immediately resuspended in 1 mL SOC and recovered by shaking for 2 hours at 37°C. 100 µL of recovered culture was plated on LB agar with kanamycin to select for recombinants.

Confirmation of *tldD* disruption was performed by streaking recombinants to single colonies on LB agar with chloramphenicol and incubating overnight at 37°C. PCR was performed on

overnight cultures inoculated from single colonies using primers flanking the recombination site to verify insertion of the kanamycin cassette. Loss of the pKD46 plasmid in recombinants was confirmed by streaking on LB agar with ampicillin at 37°C.

Growth Curve Analysis

The growth of bacterial cultures was monitored by measuring the optical density and performing viable counts over time. Optical density was measured using 11 mL of culture in 1.5 cm diameter glass tubes with a Jenway 7300 spectrophotometer at 600 nm (A_{600}). Glass tubes were used instead of cuvettes to minimize contamination during repeated measurements. An A_{600} of 0.4 corresponds to approximately 6 x 10^7 cells/mL.

For viable counts, serial dilutions of the culture were made in 1 mL of M9 minimal medium, ranging from 10^0 to 10^-7 dilutions. 10 µL of the 10^{-5} to 10^{-7} dilutions were spotted in duplicate onto LB agar plates. After incubation, colonies were counted for the spots containing between 10-30 colonies, as higher densities made individual colonies difficult to discern. The viable cell titre was calculated from the dilution factor and spotted volume using the Viable Titre function in R.

The growth curve experiment was performed by inoculating fresh medium and measuring the A_{600} every 30 minutes. At each timepoint, viable counts were also determined as described above. The viable titres were plotted over time using ggplot2 in R. This procedure allows quantification of growth kinetics and viable cell number over time. Repeated measurements of A_{600} and viable counts enables comparison of growth between different conditions or bacterial strains.

Spot Dilution Assay

To evaluate growth characteristics and viability, bacterial strains were analyzed by spot dilution assay. Cultures were grown to a defined optical density, then serially diluted down to 10^-7. 10 µL of each dilution was spotted onto LB agar plates, including a no dilution spot, and plates were incubated overnight. Viable titres were calculated taking into account the initial culture volume, dilutions, and volume plated using dilution mathematics. Growth was assessed by analyzing the last dilution showing growth. This provided quantitative viable titres and qualitative analysis of growth phenotypes. The assay was performed in biological replicates and across a time course

to generate growth curves. Spotting standardized volumes enabled quantitative comparisons between strains. The serial dilutions provided visual determination of viability down to single colonies.

Phylogenetic Analysis

To elucidate the conservation of the replication fork trap architecture in *E. coli*, I first identified representative strains from major phylogenetic groups including A, B1, B2, D and E. Complete genomic sequences for these strains were retrieved from the NCBI database to serve as reference genomes.

To precisely map the locations of genes and DNA sequences, I leveraged the sensitivity of the Bowtie2 aligner. For full genes like *tus*, I generated custom BLAST databases for each reference genome using Python scripts. I then executed local BLAST searches to pinpoint gene coordinates on the chromosomes. For smaller sequences like the 23 bp ter sites, I directly aligned them to the references with Bowtie2 using the Rbowtie2 Bioconductor package. This approach enabled nucleotide-level resolution in mapping *ter* sites.

With locations established, I next focused on alignments to assess sequence conservation. For genes, I utilized the robust multiple sequence alignment algorithms in the R packages MSA and Biostrings. For *ter* sites, I could extract the aligned sequences from the Bowtie2 output and analyze conservation, particularly of the crucial GC6 bp. Potential *ter* sites were excluded based the consensus conserved inner sequence for ter sites and if the C6 base pair was missing or if other Tus binding bases were mismatched to a threshold of 5.

Phylogenetic analysis provided evolutionary context on the conservation data. I generated a Tus protein tree by aligning sequences with MSA and then constructing a distance matrix with the Phylotree package. The tree revealed striking maintenance of Tus across niches.

This bioinformatic workflow provided a multilayered view into fork trap architecture conservation. By leveraging mapping, alignment and phylogenetic approaches, I could relate conservation of the system to evolutionary trajectories of *E. coli* strains. The methodology yielded key insights that advance our understanding of this bacterial replication control mechanism. I anticipate this workflow can be broadly applied to elucidate conservation in other genomic systems.

Recombination Rates Analysis

Mutation rates were measured using a Luria-Delbrück fluctuation test as described previously (Swings et al., 2017). Briefly, overnight cultures of the *E. coli* strains were diluted 1:100 into 1 mL of Mu medium in 2 mL tubes to achieve an initial A_{600} of 0.04. For each strain, 11 parallel

cultures were grown at 37°C with shaking at 1000rpm in a thermomixer until the OD reached an A_{600} of 0.4. One additional small culture was grown alongside the fluctuation test cultures to measure A_{600} in a cuvette for determining the viable cell titre. Viable titres were determined by spotting serial dilutions of the parallel A_{600} culture onto agar plates three times and the average colony count was used to represent the number of colonies for that dilution. Dilutions of 10e-5 and 10e-6 were used to avoid resolution issues for higher dilutions. When the target A_{600} was reached, cultures were centrifuged at 6000 x g for 5 min and resuspended in 100uL M9 minimal medium. The parallel test cultures were plated onto Mu agar supplemented with 40 µg/mL kanamycin to select for mutants in which a reversion had taken place. Plates were incubated at 37°C for 24 hours until colonies formed. Colonies were counted by thresholding images in ImageJ.

Frequencies were calculated by mutants/total cells and mutation rates were calculated from colony counts using the Flan R package (Mazoyer et al., 2017), implementing the Ma-Sandri-Sarkar maximum likelihood estimator generally accepted as the preferential method for determining rates from fluctuation data (Gillet-Markowska et al., 2015; Zheng, 2017). By plating whole cultures this method avoids dilution and pipetting errors as the total number of colonies counted represents the total number of mutants in the 1mL culture. Unlike standard mutation rate analysis where cultures are grown to saturation, this method use cultures grown to a defined OD as the reversion rates using this system are orders of magnitude higher than that of standard mutation rates commonly used (Foster, 2006; Luria and Delbrück, 1943).

R-loop Detection by Dot Blot

R-loop formation was analyzed by dot blotting as described previously (Raghunathan et al., 2019; Ramirez et al., 2022; Vlachos-Breton and Drolet, 2022). Overnight cultures were grown in Mu minimal medium and used to inoculate 11 mL subcultures to an A_{600} of ~0.04. Cultures were grown to A_{600} 0.4, centrifuged at 5000 x g for 5 min, and resuspended in ice-cold PBS. Genomic DNA was extracted using the Monarch Genomic DNA Purification Kit (NEB #T3010) and quantitated on a NanoDrop spectrophotometer. Samples were diluted to 20 ng/µL in elution buffer.

For dot blotting, 5 µL of DNA was spotted onto Hybond nitrocellulose membrane (GE Healthcare). DNA was UV crosslinked to the membrane using a UV stratalinker 1800 (Stratagene) at 120 mJ/cm^2. Membranes were blocked for 1 hr at room temperature with 2% non-fat milk in

PBS. Blocked membranes were incubated for 2 hr at room temperature with the S9.6 antibody (Kerafast) diluted 1:5000 in 0.1% PBST, followed by 5 x 5 min washes with 0.1% PBST. Membranes were then incubated for 1 hr at room temperature with HRP-conjugated goat anti-mouse secondary antibody diluted 1:5000 in 0.1% PBST, followed by 6 x 5 min washes. Chemiluminescent detection was performed by incubating membranes for 5 min with ECL reagent (BioRad) at room temperature and imaging on a G:Box Chemi XX6 (Syngene) with a 30 sec exposure.

Dot intensities were quantified using ImageJ software. Background subtraction was performed by measuring intensity of a blank region of membrane and subtracting this from each dot intensity value. R-loop levels were normalized to the level in wild-type cells. Data analysis and plotting was conducted in R using ggplot2.

Agarose Gel Electrophoresis

To analyze and resolve DNA fragments, agarose gel electrophoresis was performed. A 1% agarose gel matrix was prepared by dissolving agarose powder in 1x TBE buffer through heating. The gel solution was cooled and cast in a mould to polymerise. DNA samples were mixed with loading dye and loaded into the wells alongside a molecular weight ladder. The gel was run at 120V for 40 minutes to achieve adequate size-based separation of fragments. After electrophoresis, the gel was stained with SYBR safe and DNA bands were visualized under UV light. This standard protocol enabled analysis and verification of PCR products and other DNA fragments by exploiting the size-dependent mobility of nucleic acids through agarose gels. Key parameters like percentage, voltage and runtime were optimized for resolution of fragments in the desired size range.

PCR

Polymerase chain reaction (PCR) was utilized to amplify target DNA sequences. For recombineering applications, Velocity DNA Polymerase from Bioline was used due to its high fidelity and processivity. PCR reactions contained Velocity buffer, dNTPs, primers designed with 40-60% GC content, DNA template, and polymerase. Thermocycling conditions included an initial denaturation at 98°C; followed by 25-35 cycles of denaturation at 98°C, annealing at primer Tm, and extension at 72°C; final extension was performed at 72°C. For verification of constructs, OneTaq DNA Polymerase from NEB was used. OneTaq reactions contained standardized buffer, dNTPs, template, primers, and polymerase. Cycling conditions were initial denaturation at 94°C; followed by 25-30 cycles of denaturation at 94°C, annealing at primer Tm, and extension at 68°C; final extension was at 68°C. Appropriate polymerases and optimized reaction conditions enabled efficient, specific amplification of target sequences for downstream applications.

gDNA Extractions

To obtain high quality genomic DNA (gDNA) from bacterial samples, the Monarch Genomic DNA Purification Kit (NEB #T3010) was utilized per the manufacturer's protocol. Briefly, cell pellets were resuspended in lysis buffer containing RNase A to degrade RNA. Proteins were precipitated and removed using a protein precipitation buffer. gDNA was then bound to a DNA binding

buffer, washed, and eluted in TE buffer. Extracted gDNA was quantified by NanoDrop and normalized to a concentration of 20 ng/µL for downstream applications. This kit-based protocol provided an efficient, standardized approach to purify gDNA while avoiding use of hazardous organic extractions. By following the optimized kit procedure, high molecular weight gDNA free of proteins and RNA was obtained at sufficient yields and purity for techniques such as PCR, sequencing, and genomic library preparation. Normalization to a defined concentration enabled direct use of the gDNA in sensitive enzymatic applications.

DNA extraction from agarose gel

DNA extraction from agarose gels The bands containing the DNA required were purified using the NucleoSpin® Gel and PCR CleanUp kit (Macherey-Nagel) according to the manufacturer's instructions. Briefly, the plugs cut out of a TAE gel were mixed with a chaotrophic salt-containing binding buffer and incubated at 50°C until the agarose gel plugs dissolved. The sample was loaded on to a NucleoSpin® Gel and PCR Clean-up Column where, in the presence of chaotrophic salt in the binding buffer, the DNA binds to the silica membrane. A number of washes were carried out with ethanolic buffer to remove any impurities and finally the DNA was released from the silica by eluting with a low ionic salt solution. The sample was measured using a Qubit.

Software and Data Analysis

Genomes for bioinformatics chapter

All *E. coli* and *Shigella* genomic sequence were downloaded from the NCBI nucleotide database.

Table 5: NCBI E. coli strains used for bioinformatics analysis of ter sites and replication fork trap architecture

Organism	Strain	Phylogroup	Accession
E. coli	MG1655	A	NC_000913.3
	BW2952	A	CP001396
	REL606	A	CP000819.1
	APEC078	B1	NC_020163.1
	IAI1	B1	NC_011741.1
	11368	B1	NC_013361.1
	S88	B2	NC_011742.1
	UTI89	B2	NC_007946.1
	E2348	B2	NC_011601.1
	UMN026	D	NC_011751.1
	IAI39	D	NC_011750.1
	SMS-3-5	D	NC_010498.1

CE10	D	NC_017646.1
042	D	NZ_CP042934.2
TW14359	E	NC_013008.1
Sakai	E	NC_002695.2
EDL933	E	NC_002655.2

Table 6: NCBI *Shigella* strains used for bioinformatics analysis of ter sites and replication fork trap architecture

Organism	Strain	Accession
Shigella	flexneri 2a str. 301	AE005674.2
	flexneri Shi06HN006	CP004057.1
	flexneri 5 str. 8401	CP000266.1
	flexneri 2002017	CP001383.1
	flexneri 2003036	CP004056.1
	boydii CDC 3083-94	CP001063.1
	boydii 600080	CP049606.1
	boydii 600090	CP049278.1
	boydii Sb227	CP000036.1
	dysenteriae Sd197	NC_007606.1
	sonnei Ss046	CP000038.1

Phylogenetic analysis

To analyze conservation of the replication fork trap architecture, I leveraged a multifaceted bioinformatics toolkit. Genome sequences were retrieved from NCBI to serve as references. Precise mapping of genes and DNA sites was achieved using BLAST and Bowtie2 alignment driven by Python and R scripts. Conservation was assessed by aligning sequences with R packages like MSA and visualizing phylogenetics with Phylotree. The tidyverse enabled key data manipulation and figure generation. Molecular modeling in ChimeraX provided structural insights into protein-DNA interactions. AlphaFold2 predicted mutant Tus conformations, revealing localized impacts on the ter binding pocket. Custom Python GUIs increased workflow efficiency. CDS analysis was performed by Prodigal and parsed using BLAST and Python scripts. tBLASTn aligned Tus proteins between species.

This integrated toolkit combining genomic mapping, alignment, phylogenetics, modelling, prediction, visualization and automation enabled in-depth characterization of fork trap conservation across diverse *E. coli* genomes. Rbowtie2, MSA and other packages provided critical capabilities for sequence analysis. ChimeraX and AlphaFold2 yielded key structural insights. Custom scripting increased throughput. Together these computational resources elucidated the evolutionary maintenance of the replication fork trap architecture in bacteria, advancing our understanding of this conserved replication control mechanism.

Machine learning structural predictions of R-loops and G4s

G-quadruplexes (G4s) were predicted using the machine learning tool G4Boost run in the command line on MG1655 reference genome, using a more restrictive approach to determine G4 forming sequences (Cagirici et al., 2022). Command: *python G4Boost_v4.py –fasta MG1655.fasta –maxloop 10 –minloop 1 –maxG 4 –minG 3 –loops 4*. See GitHub page for more details on what each parameter represents (GitHub - hbusra/G4Boost). The output gff file was parsed through a custom R script to convert into csv format, cleaning for only the position and length of the potential G4 site.

R-loop forming sequences were predicted from the MG1655 reference genome using the webtool R-looptracker (Brázda et al., 2021). Available at: DNA analyser (ibp.cz). Raw csv output was downloaded following analysis and cleaned using a custom R script retrieving position and length information.

Both the R-loop and G4 data was piped into the ForkTrapApp (custom Python app developed by myself) to visualise custom genomic features in the context of the *E. coli* genome and the replication fork trap. Separate tracks were selected for the R-loop and G4 positions for clarity. R-loops coloured red and G4s coloured green.

R functions for Viable Titre

For calculating the doubling time and viable titre, an in house R function was created to input the colony counts then output the viable titre information to a table format. Doubling times were calculated with the linear modelling function lm() in R. See example of the function and output below. The viable titer calculations involve dividing the sum of the average colony counts by the product of a fixed factor (0.1) and the sum of the dilution factor. This formula takes into account the dilution factor and normalizes the colony counts to obtain the viable titer values.

Table 7: Tables from each time point were concatenated into a combined dataset then parsed through the ggplot2 library to produce the growth curves seen in figures 7-9

time	dilution	WTcoloniesA	WTcoloniesB	MUTcoloniesA	MUTcoloniesB	WT_VT	tus_VT
0	0.001	2	2	1	2	15909.09	11363.64
0	0.010	10	21	11	11	15909.09	11363.64

Chapter 1: Phylogenetic Analysis of the Replication Fork Trap Architecture

In *Escherichia coli*, the replication fork trap (RFT) system comprised of Tus protein binding to genomic *ter* sites which funnels fork progression and fusion during the final stages of replication into the termination area. With Tus-*ter* binding in a polar manner, forks are able to enter the fork trap, but are unable to leave. The 10 *ter* sites in *E. coli* span ~45% of the chromosome, and recent ChIP-Seq data show that only the inner most 6 sites, *terA-E* and *G* are bound by Tus protein and therefore actively arrest forks, forming the primary replication fork trap (Toft et al., 2021), with all other *ter* sites now being considered as secondary, in line with earlier 2D electrophoresis findings (Duggin and Bell, 2009).

Ideally, forks will enter the fork trap at roughly the same time and fuse, whereby chromosomal dimers are formed which will then need decatenation by XerCD recombinase using the two copied of the *dif* sites (Barre, 2007; Bigot et al., 2005; Massey et al., 2004). However, if one fork is held up on its way to the termination area due to DNA damage, lesions or collisions with transcribing RNA polymerases, then the fork trap allows the opposite fork to be stably bound at Tus-*ter* barriers. When the cause of the delayed replisome has been overcome, then it can meet with the stably bound replisome which is held up in the termination area, whereby fusion can take place. Without the fork trap present in this context of DNA damage, forks would continue to replicate the chromosome past the termination area, where fork fusion sites will be more stochastic.

Naturally, the main purpose of the fork trap system is allow forks to fuse in a defined area, reducing variability of this position and allowing the cell to deal with any consequences of fork fusions by constraining the event to a specific location (Dimude et al., 2016; Rudolph et al., 2013). However, this does not adequately explain why this system has evolved in *E. coli*, a gram negative bacterium, and a similar system in *B. subtilis*, a gram positive bacterium, but not in all prokary-

otes. Many roles have been suggested for the function of the RFT, yet the exact purpose and evolutionary driver has not been fully elucidated. There is research that shows the RFT helps to restrict pathological over-replication occurring during termination, caused by fork fusions, but it has also been shown that the RFT plays a role in coordinating termination with cell division and segregation of sister chromosomes (Galli et al., 2019; Lemon and Grossman, 2001; Mercier et al., 2008).

Most previous bioinformatic data has been performed in MG1655 or W3110, which even though there is an inversion at *oriC* can also be considered a wild type strain. Early analysis by Duggin and Bell categorized *ter* sites in *E. coli*, leading to the biochemistry data in 2015 by groups such as Berghuis, Pandey and Elshenawy (Berghuis et al., 2015; Elshenawy et al., 2015; Pandey et al., 2015), being summarized in the 2018 review (Berghuis et al., 2018) bringing the consensus of Tus-*ter* binding and replication fork arrest by different *ter* sites. As most of the *ter* site and replication fork trap architecture has been studied in MG1655, we can ask the question how conserved is the fork trap across different *E. coli* lineages occupying diverse niches, as well as related bacterial species? One might expect there to be considerable variability between genomes of *E. coli* which occupy different niches, especially if these strains are pathogenic or commensal.

Works by Touchon and Abram classifying *E. coli* phylogenetic groups classified all known *E. coli* and *Shigella* genomes into distinct groups, which I have used to match MG1655 *ter* sites and their sequence conservation throughout the phylogenetic groups (Abram et al., 2021; Touchon et al., 2009) . This analysis was performed using groups; A, B1, B2, D, E from *E. coli* and S from *Shigella*, these included strains which are more ancestral, such as B2, and those which have more recently emerged as independent groups, such as group A. More recent phylogenetic analysis has since characterized groups F and G, which house some distinctly pathogenic strains, and a further analysis was performed taking representative genomes from these groups, to fully include the possible variations in the replication fork trap system within *E. coli* (Abram et al., 2021; Clermont et al., 2019). The work from (Goodall et al., 2021) did not include group F and G, this analysis was performed after publication.

This chapter investigates the fork trap through bioinformatic analysis of *ter* site conservation across the Enterobacteriaceae and sheds light on the importance of maintaining an active RFT across diverse *E. coli* strains. This aims to advance understanding of this key bacterial replication control mechanism by delineating specific *ter* sites critical for arresting replication, and elucidating why the RFT system has been evolutionarily conserved.

The aims of this chapter are to identify growth differences between wilt type and *tus* single mutants, to analyze the sequence conservation of *ter* sites across a variety of bacterial species including and related to *E. coli*. To view the overall architecture of the fork trap system to identify any consistent patterns in the location of specific *ter* sites relative to each other, the arithmetic midpoint and the *dif* dimer resolution site. To identify the structural differences of known Tus mutants to further elucidate how structure of Tus and its binding to DNA is related to its function (Berghuis et al., 2018; Kamada et al., 1996; Mulcair et al., 2006).

Together, these data build on from previous research and further deepen our understanding of the replication fork trap system in bacteria and provide insight into its overall evolutionary conservation. Paired with the other chapters in this thesis, I present a strong case that the fork trap is an important and beneficial system for *E. coli*, and related species, that helps us understand more about mechanisms of replication termination in bacteria (Goodall et al., 2021, 2023).

Tus and RFT inactivation:

Previous research showed that cells which have a *tus* inactivation, and therefore lack an active replication fork trap, have no phenotype and are able to grow as well as wild type, opening up the question of the purpose of this genomic architectural system (Roecklein et al., 1991). As more modern sequence analysis tools were developed, more recent data has been gathered that suggest *tus* mutants have a mild growth defect of around 1 minute Goodall et al. (2021). To test if there were any growth difference between the wild type and *tus* single mutant, cells were grown in LB and incubated at $37^{\circ}C$ to an A_{600} of 0.4 then serial diluted where a spot dilution assay was performed for each 30 minute time point for a total of 3 hours (see methods for further details).

I saw a consistent trend across the data sets that *tus* cells had between 0.5-1 minute growth delays that averaged out to around 0.5 minutes difference, Figure 18 below shows the linear growth curve of wild type MG1655 compared to *tus* single mutants. Being that the difference in growth is not that extreme, representation of these viable titres over time is better represented as in Figure 18.

From these growth curve data sets we can see that the average doubling time for WT is 19.1 minutes whereas the *tus* mutant has an average doubling time of 19.6 minutes (Figure 18). This difference, however small, was consistent throughout the experiments and we conclude that at worst the *tus* cells grew slightly slower than the wild type cells. This essentially means that, *in vivo*, replication forks will proceed as usual once initiated from *oriC* and continue until they fuse opposite *oriC* , this time there will be no control of forks moving outside of the termination area.

This reiterates the question of why certain bacteria have a replication fork trap, such as *E. coli* and *B. subtilis*, yet the majority do not need such a system (Galli et al., 2019). Rudolph and colleagues have shown through marker frequency analysis that *tus* mutants have a less pronounced valley in copy number in the termination area, showing that forks are permitted to move out of the termination area when Tus is not present, on occasion, but without Tus there is more variation in the position of fork fusion events, which will present itself as a flatter valley in the MFA data (Dimude et al., 2016; Rudolph et al., 2013).

From this work, I decided to perform bioinformatics analysis on a variety of *E. coli* strains which cover the phylogeny of this species, to observe any trends in the variation of this sequence dependent system. If a replication fork trap results in only mild growth disadvantage then we might expect there to be significant variation in the chromosomal architecture within the ter-

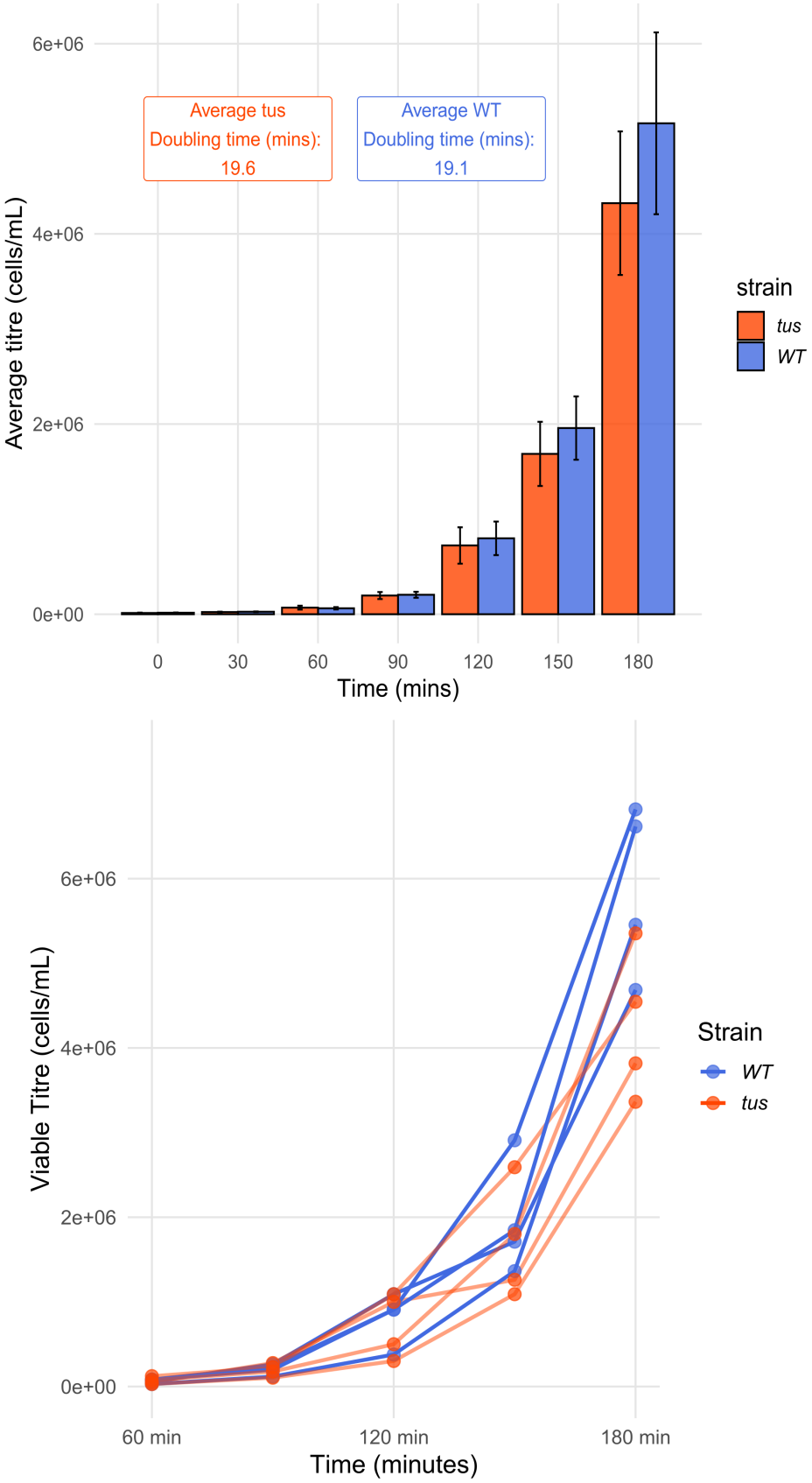

Figure 18: Linear growth curve of WT vs *tus* mutants. Growth curve averages from 4 independent datasets. Error bars are calculated from the SEM. Representation in bar graph format compared with the standard line graph format to more easily visualise the subtle differences in viable titre for each data point, for visual clarity, line graph begins at 60 minutes. A Wilcoxon test showed significant different between WT and tus mutants across the 4 datasets.

mination area with strains which experience select genomic pressures from their environment. With a broad initial hypothesis of simply measuring this conservation, we might expect to see changes in the order and orientation of *ter* sites, variations in the sequences of the 23 bp *ter* sites and possibly even the amino acids that make up Tus protein would be altered show different sequence altogether.

The Replication Fork Trap Across Phylogenetic Groups

The ten *ter* sites (*A-J*) were mapped to all *E. coli* genomes and matched as a correct *ter* site through application of a rigid exclusion criteria. The most important of which was ensuring the crucial GC6 base pairing was present, this has been studied extensively to be necessary for the 'molecular mouse trap' to be activated. If any single nucleotide change was present implicated in Tus binding to the bases themselves then the site was also excluded. To perform a general filter following the bowtie2 alignment, the mismatch threshold need not exceed 5 when compared to the MG1655 *ter* sites and that most variability seen was in the outer *ter* region, leaving the inner core of the *ter* site mostly intact. Figure 19 shows the architecture of the matched *ter* sites that then map to the genome forming the fork trap system.

E. coli populations and strains can be quite diverse, given the difference in niche and virulence, with a core genome of roughly 2000 genes and pangenome of 13,000 there is clearly some variability across the species. The main phylogroups of *E. coli* have been well characterised as A, B1, B2, D, E and S, with new classifications being defined in more recent research, (G and F), as in Figure 3. Commensal strains are mostly inside group A, pathogenic strains in group B2 and D, then the enterohemorrhagic O157:H7 strain are in group E, with *Shigella* filling the S group. In K12, the inner *ter* sites *A-E* are not in ORFs, whereas the outer *ter* sites are present within a gene and we might expect the conservation of outer *ter* sites to be mainly from the genes in which they reside, and not necessarily due to their ability to block forks.

Even so, the variability of *ter* sites *A-E* should give us an indication of much they contribute to the fitness of *E. coli*. (Duggin and Bell, 2009) showed that few forks are actually paused at the outer *ter* sites and could be labelled as pseudo *ter* sites and this idea has been further developed in my own work (Goodall et al., 2021) and more recently in (Toft et al., 2021).

As the effect on doubling time is not that dramatic in cells lacking a fork trap, we might expect variability in the arrangement of *ter* sites throughout the *E. coli* phylogroups, seeing as variability would come at little cost. The analysis began with a single genome from groups A, B1, B2, D and E, using MG1655 as the template at which to compare all other results, seeing as the majority of the previous analysis has been performed in this strain.

The initial analysis showed variability in the genome sizes of each strain, but all 10 *ter* sites could be identified with little trouble. The inner *ter* sites *A-D* also showed a relatively consistent span of the genome, 4-6%, which is also seen in MG1655. To account for the variability

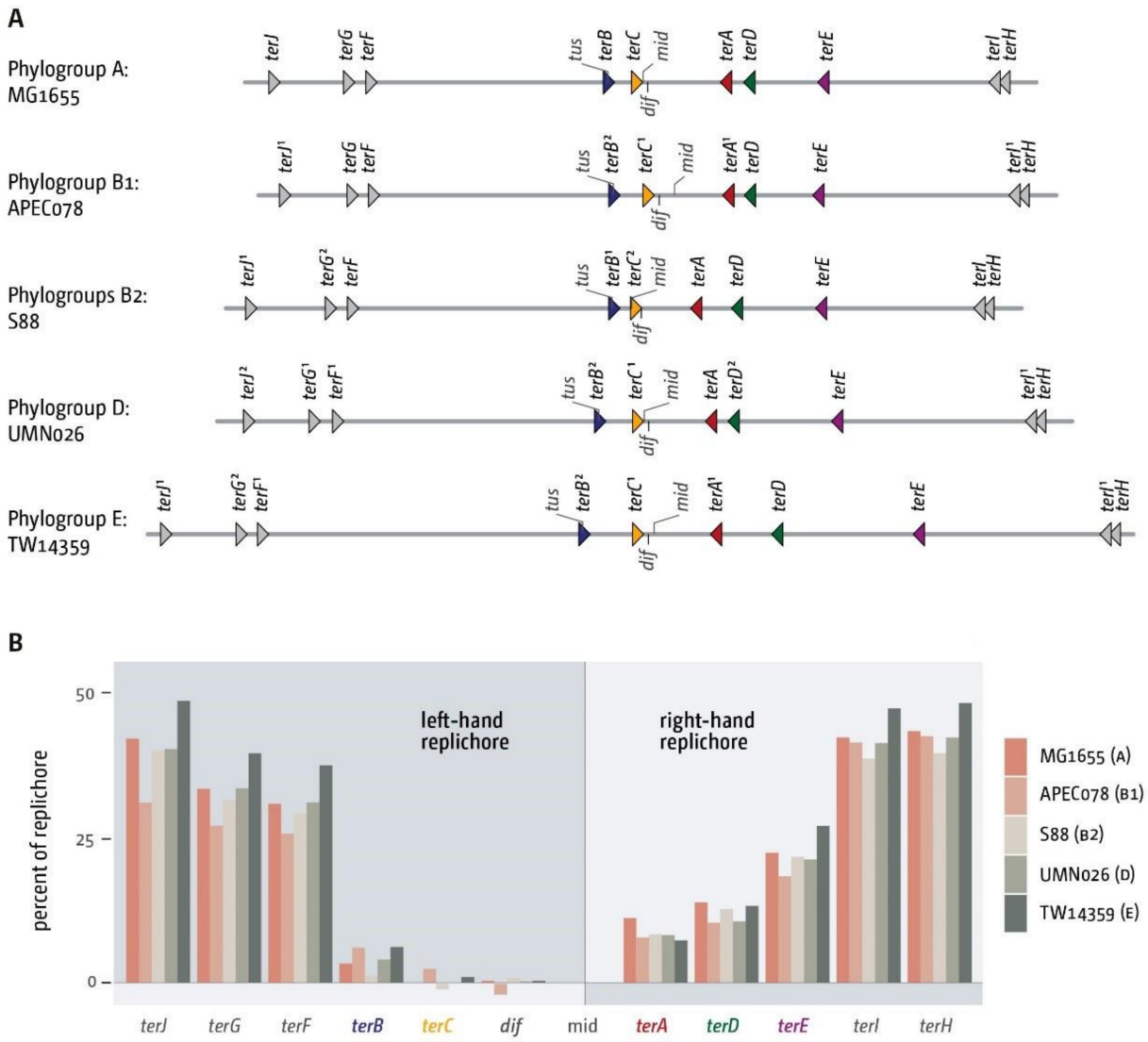

Figure 19: Architecture and genomic distribution of replication fork trap sites in *E. coli* genomes across major phylogroups. (A) Absolute positions of all primary *ter* sites in *E. coli* genomes representing phylogroups A, B1, B2, D, and E. All *ter* sites shown as triangles, with orientation indicating blocking direction. Chromosome dimer resolution site *dif* marked by tick. Single nucleotide changes versus MG1655 indicated by superscript. *dif* strand location shown by marker direction. (B) Comparative analysis of relative *ter* site positions within chromosomal context across phylogroup genomes. To normalize for genome size differences, theoretical midpoint calculated for each genome as reference. *ter* and *dif* positions calculated as percentage of replichore length (*oriC* = 100%). Indicates variability in relative positions of these elements between genomes. Phylogroups denoted in brackets after strain names.

in genome sizes, we calculated the mid point based on *oriC* and the calculated the percentage of the replichore that needed to be replicated before a fork encounters a specific *ter* site. These calculations showed that for 4/5 genomes the mid-point was located between the inner *ter* sites *A* and *C*, close to the *dif* site (Figure 19B). Only in S88 (group B2) was the mid-point outside of *terC*, between *terC* and *terB*. Seeing as how most forks do not proceed at the same rates in perfect symmetry, it is not clear without experimental work if forks regularly fuse at *terC* in S88.

The mild positional variations on the chromosome when compared with MG1655, but the order of *ter* sites remained the same. We analysed the sequence variations of all primary *ter* sites, in which we saw some variability between 1-2 bp, but these variations were mostly restricted to the outer area of the *ter* sites, which do not possess direct interactions with Tus protein and are therefore unlikely to affect binding. Position 7 routinely showed some variability, which could be concerning as this is in the inner sequence, but biochemical analysis has shown this base does not interfere with Tus and is allowed to have some variability without compromising Tus binding and RFT integrity.

Only UMN026 from group D showed a variation in the GC6 base for *terJ*, where the guanine had been substituted for an adenine. We could therefore hypothesize that *terJ* in UMN026 would not be as effective at pausing replication fork progression, based on this *in silico* result, where molecular analysis would need to be performed in future for confirmation. In 180 analysed *ter* sites, across 18 genomes, this was the only exception for a *ter* site in which we could theorise would have reduced Tus binding affinity from likely not being able to form the 'mouse trap' lock in crucial residues, as stated in (Mulcair et al., 2006).

Naturally, without performing biochemical analysis on this *ter* site variation we cannot be certain of how much this would impact Tus binding, however from works by Elshenawy and Berghuis we can suggest that there may be reduced replication fork blocking in this strain's *ter* site (Berghuis et al., 2015; Elshenawy et al., 2015)

In terms of how likely this is to impact the overall functionality of the replication for trap in UMN026, *terJ* is rarely used a site which forks are paused at and so there is little to suggest this change would have impacted the functionality of the replication for trap enough to warrant correction. Also, *terJ* being located within an ORF is more likely to have the the change as a result of gene mutation that the *ter* site directly as this is has the greater selection pressure.

Tus biochemistry

Tus alignments

My analysis of ter site conservation across *E. coli* phylogenetic groups revealed strong maintenance of the fork trap architecture. I therefore hypothesized the Tus protein would also show conservation, particularly of residues involved in *ter* binding. This would maintain the specific interactions underlying formation of the Tus-*ter* complex that enables replication fork arrest.

To test this prediction, I aligned Tus proteins from diverse *E. coli* strains (Figure 20). Residues contacting the *ter* DNA backbone and bases are highly conserved. In particular, E49 which forms crucial hydrogen bonds with the GC base pair at position 6 is identical across all Tus homologs analyzed. This lock is critical for the sequence-specific interaction enabling the trap to arrest replication forks. All other residues which contact the *ter* site were also maintained, where variability is highlighted it is not at a position which should alter the efficiency of Tus binding to *ter* DNA (Figure 20).

The strict conservation of key Tus residues provides a molecular rationale for the maintenance of a functional fork trap across niches. *E. coli* has faced various selection pressures as strains adapted to different environments. Yet the fork trap has been uniformly preserved, likely due to the importance of regulating DNA replication. The conservation of Tus residues that recognize ter sites and arrest fork progression reveals how the system has been evolutionarily maintained at the molecular level.

Overall, by correlating ter site and Tus conservation, these findings strengthen the model that the fork trap architecture is universal in *E. coli*, not just limited to laboratory strains. The system appears to provide an important replication control function that has prevented variability across diverse lineages. My analysis provides both bioinformatic and biochemical explanations for the evolutionary maintenance of this arrest system in bacteria.

Structural prediction of Tus Mutants

The Tus mutant data is clear on the features of each mutant and their probability of Tus-*ter* lock formation and ability to arrest replisomes. We understand that specific interactions are needed to mediate the Tus-*ter* C6 lock, but also generalized stability of the protein-DNA complex.

By using Deep Mind's Alpha Fold 2 algorithm I predicted the structure of Tus and show the struc-

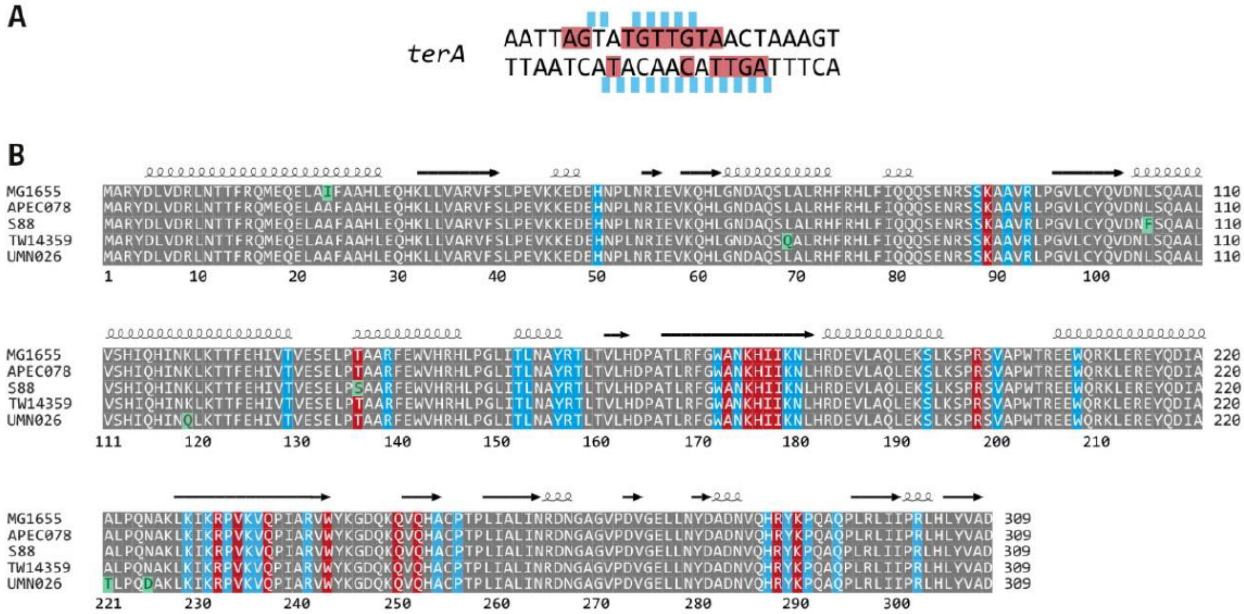

Figure 20: Interactions between Tus protein and ter sequences. (A) *terA* sequence in which interactions of Tus protein with the DNA backbone or individual bases is highlighted, as previously described. Backbone contacts are shown in light blue above the terA sequence, where bases in direct contact with Tus are highlighted in red. (B) Sequence alignment of *E. coli* Tus proteins from the 5 phylogenetic groups. Identical residues are shaded grey, whereas any individual changes are highlighted in green. Those residues that interact with the DNA backbone of the terA sequence are highlighted by a blue block, while residues that make sequence-specific contacts with the *terA* DNA are highlighted with a red block. Secondary structure information provided above is based on analysis using the ENDscript 2 server: Taken from Goodall et al.,2021

ture to be near identical to the crystal structure. This application could be used to decipher the impact Tus mutations would have on the overall structure of Tus, as discussed in previous research. Tus mutants that maintain C(6) interactions include Q250A, R198A and E49A, whereas those that do not include F140A, H144A and E49K (Table 8).

Table 8: Key amino acid residues of E. coli Tus protein and their role in blocking progression of replication fork complexes.

Tus_residue	Importance	Reference
E47	Stabilize 'mousetrap'	*Mulugu et al. (2001) and Toft et al. (2021)*
E49	Slow approaching fork via helicase interaction	*Berghuis etal., (2015)*
F140	Holding C6 in place	*Mulcair et al., (2006)*
H144	Holding C6 in place	*Mulcair et al., (2006)*
G149	Holding C6 in place	*Mulcair et al., (2006)*
R198	Slow approaching fork via helicase interaction	*Elshenawy et al., (2015)*

Showing structure changes in these predictions and noting any drastic structural changes in Tus from the WT that could provide tertiary conformation change information that may also likely contribute to lack of Tus binding and C6 lock formation.

The mutants in question all centre around the C6 binding pocket forming the so-called mouse trap. The AlphaFold2 data presented here shows local conformation changes compared to the wild type Tus protein providing structural insight into the lack of functionality in these mutants.

The overall structure of Tus in each of these mutants is not significantly disrupted, although small changes are apparent, illustrating it is local conformation changes in the binding pocket itself which are disrupting the ability of Tus to form Tus-*ter* locked complex, in some mutants it has been calculated that the locked complex can still be formed albeit with reduced probability.

Being able to predict DNA binding affinity of varying mutants, not just for Tus but all DNA binding proteins, will aide our understanding of how these interactions are formed.

Recent advances in protein structure prediction algorithms such as AlphaFold2 provide new opportunities to gain structural insights into the impact of mutations on protein conformation. I utilized AlphaFold2 to predict the structure of several Tus mutants that have previously been characterized functionally (Berghuis et al. 2015; Elshenawy et al. 2015; Pandey et al. 2015). The mutants analyzed included E49K, F140A, H144A, R198A, and Q250A (Figure 21).

The AlphaFold2 models reveal that the overall structure of Tus is largely maintained in these mutants. However, localized conformational changes are apparent within the C6 binding pocket, providing a structural rationale for the observed functional effects.

Specifically, mutants F140A and H144A are predicted to disrupt the "lock" interaction with C6 of the DNA at the *ter* site, correlating with biochemical data showing these mutants abolish arrest activity. In contrast, mutants such as Q250A and R198A retain the potential to lock C6 in the correct orientation. However, even in these mutants, the conformation of the binding pocket is altered compared to wild-type Tus, rationalizing the reduced locking probability observed experimentally.

Overall, this AlphaFold analysis provides unique structural insights into the conformational effects of mutations on the key C6 binding pocket of Tus. The results highlight the utility of modern protein structure prediction methods for gleaning atomic-level information to explain the impacts of mutations. Applying AlphaFold systematically to characterize panels of Tus mutants could aid in elucidating key structure-function relationships and binding determinants. Additionally, the approach demonstrated here could be widely applied to other DNA-binding proteins to better understand sequence-structure-function maps.

From Figure 21, we can see noticeable changes in the arrangement of R198, for example, in both F140A and H144A mutants. The R198 seems to be covered up, indicating a steric interaction between both F140 and H144. When either of these residues is replaced with alanine, the normal protruding R198 is no longer in the WT position. This may also lead to less effective C6 lock formation in these mutants, where previous research has simply implied that each residue acts almost independently in forming the C6 pocket.

Interestingly, the only mutant with an intact R198 placement is the crucial E49 residue. This suggests that this residue in particular may not be involved in guiding the single strand *ter*-DNA, and instead is primarily involved in holding the C6 base in place, rather than determining the overall structure of the binding pocket itself. This is in line with the data from 2015 (Berghuis et al., 2015; Elshenawy et al., 2015; Pandey et al., 2015).

Logically, the 3D structure of a protein is determined by the intramolecular forces that fold the protein into its functional shape. This data is the first to show the 3D structure of Tus mutants and the effect of each residue change in the local conformation of the C6 binding pocket.

Where altering residues that form the C6 pocket, there are naturally alterations in the shape of

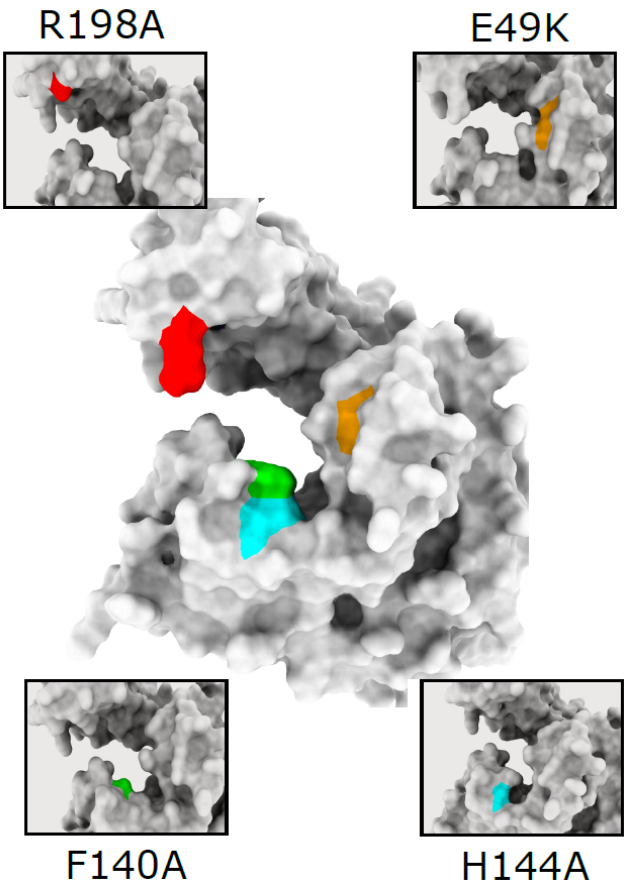

Figure 21: Comparison between WT Tus protein and the predicted structure of each notable mutant, the variation in the local conformation of the C6 binding pocket in each mutant is observable. The R198 residue is distorted in the F140A and H144A mutants, normally protruding out towards the C6 pocket. Other unique local changes are present, which are difficult to quantify, but all likely add to the defect of the mutant in binding *ter* sites and arresting forks.

the pocket itself, the fact that altering residues inside the pocket has an effect on the conformation of other residues also know to impact Tus-*ter* binding is a surprising find.

Utilizing AlphaFold2, and other machine learning models, for biochemical protein prediction can lead to future work where other mutants are identified than can give further insight into the functioning of Tus-*ter* interactions, and can be applied more broadly to other protein and enzymes involved in DNA metabolism, fork restart and repair.

Although difficult to quantify, the conformation change is noticeable to the naked eye. The idea of function being strongly linked to structure is a key concept in molecular biology and biochemistry more broadly. Noting conformation changes in mutants know to have a reduce Tus-*ter* binding efficiency, or reduced probability of forming a successful C6 lock, can aide researchers into designing new mutants to test new hypotheses and give greater insight into factors affecting replication termination.

Selection Pressure and Genetic Drift in *terY*

Duggin and Bell (Duggin and Bell, 2009) first postulated that the inner *ter* sites *A-D* could be thought of as the primary replication fork trap, showing high efficacy for Tus, whereas the outer *ter* sites E-J could be thought of as secondary *ter* sites, based off of their reduced Tus binding potential. My analysis showed that the primary *ter* sites are not within ORFs while the secondary *ter* sites are present within ORFs, or overlap to a high degree. Using Prodigal command line tool to predict the CDS of each genome we can observe which *ter* sites are within and which are not within CDS. There was a high degree of homology between the same *ter* sites across the phylogroups, likely highlighting the conservation of the gene, rather than the *ter* site itself.

One important question is how much variability in sequence and position would we expect to see in other sites, not only primary *ter* sites, if there were little selection pressures to maintain their structure. To answer this question we analysed the locations of *rrn* operons, genes coding for ribosomal proteins, tRNA genes and aminoacyl-tRNA synthetase genes. These genes have been analysed in previous works by our lab (Dimude et al., 2016), but with MG1655 as the only source of inquiry., and found they are evenly distributed across the chromosome. Here, we have shown the positions of these essential genes in all 5 phylogroups to observe the amount of variability present in sites other than those in only the termination area. We found the order of genes remains relatively consistent across the phylogroups and a there is consistently a lack of these genes inside the inner termination area.

Table 9: "Location of ter sites in or outside of open reading frames (ORFs), where " - " is outside of an ORF, " + " is inside an ORF, and " " indicates partial overlap with an ORF. Sites in bold indicate the primary and functional fork trap, those grayed out indicate lack of Tus binding as determined by ChIP-Seq data (Toft et al., 2021)"

Site	MG1655	APEC078	S88	UMN026	TW14359
terA	**-**	**-**	**~**	**~**	**~**
terB	**-**	**-**	**-**	**-**	**-**
terC	**-**	**-**	**-**	**-**	**-**
terD	**-**	**-**	**-**	**-**	**-**
terE	**-**	**-**	**-**	**-**	**-**
terF	+	+	+	+	+
terG	**+**	**+**	**+**	**+**	**+**
terH	+	+	+	+	+
terI	+	+	+	+	+
terJ	+	+	+	+	+

From Table 9 we can see that the inner *ter* sites *A-E* are not within open reading frames, while outer sites are within them. It is not surprising to therefore see a high degree of conservation in outer sites as the gene in which the *ter* sites resides would be under its own evolutionary pressure to maintain sequence. The fact that the inner *ter* sites are not within coding sequences, yet retain the same order and similar sequence across phylogroups, poses the questions of why the 23bp sequences have not been altered due to random mutation and genetic drift.

The chromosome dynamics of *E. coli* have been well studied and there is no issue with the cell rearranging it's chromosome to yield a more efficient system, and it would seem logical that if the cell can rearrange large genomic regions such as at *rrn* operons, or an entire half of the termination region, to allow unimpeded replication fork progression then it should be comparatively easy to alter a 23bp region not within an open reading frame. (Dimude et al., 2018b) showed there were GCR in *rrnCABE* in a Δ*oriC oriZ+* strain to align replication with transcription.

To analyse sites in the chromosome that could glean insight into what we would expect a *ter* site to look like if it did not participate in the RFT, we studied the position of pseudo *ter* sites

(*terK,L,Y,Z*) (Duggin and Bell, 2009). We observed that *terK* and *terL* are in the same replichore as *terC,B,F,G,J* and share the direction of these *ter* sites, meaning that would, theoretically, arrest forks approaching from the right hand replichore, yet they have very weak Tus interaction and fork stalling ability. Both *terY* and *terZ* are near the origin, but in blocking orientation of forks coming directly from *oriC* , but fork pausing activity is very low and require Tus saturation to see any notable pausing (Duggin and Bell, 2009). *TerK,L* and Z are all within ORFs, whereas *terY* is not.

In line with our other observation in essential genes, we observed little sequence alignment variation in these secondary *ter* sites (not shown). However, *terY* not being within a CDS should give us some insight into the variability we would expect to see from a short sequence with little to no selection pressure to maintain its location and sequence. The presence of *terY* revealed a highly variable site, both in terms of chromosome position and sequence across all *E. coli* genomes analysed (Figure 22). We found *terY* in 10 genomes with zero mismatches, which then has to be relaxed to 5 mismatches to score a hit in another 6 genomes. We did not observe *terY* in UMN026, of group D.

From Figure 22, we can see that groups A, B1 and B2 showed little variation in location and sequence, while groups D and E were the most viable, with many SNPs and positional changes, when compared with MG1655. These sequence changes were not specific to the outer *ter* sequence, as we have seen in primary *ter* sites, and are therefore likely to seriously affect the binding of Tus, although all *terY* sites show the conserved GC6 base pairing. Interestingly, we saw that strain 042 of group D has the same *terY* sequence as MG1655, indicating that the variability of *terY* is not locked to the phylogroup, also shown in order of *ter* site for group D.

This supports the idea that without a selection pressure imposed on *ter* sites outside an ORF, *ter* sites are likely to be more variable and at risk of becoming lost from the genome, as seen in strains requiring 5 mismatches and the lack of *terY* in UMN026. The conservation of inner primary *ter* sites *A-E* which are not within ORFs, support the idea that these sites are of importance to the cell and incur some benefit (Figure 22). Also, the conservation of secondary *ter* sites *terK,LZ* present inside ORFs indicate these are likely to be pseudo *ter* sites that are retained based via selection pressure for the associated gene, and not the *ter* site itself (Duggin et al., 2008).

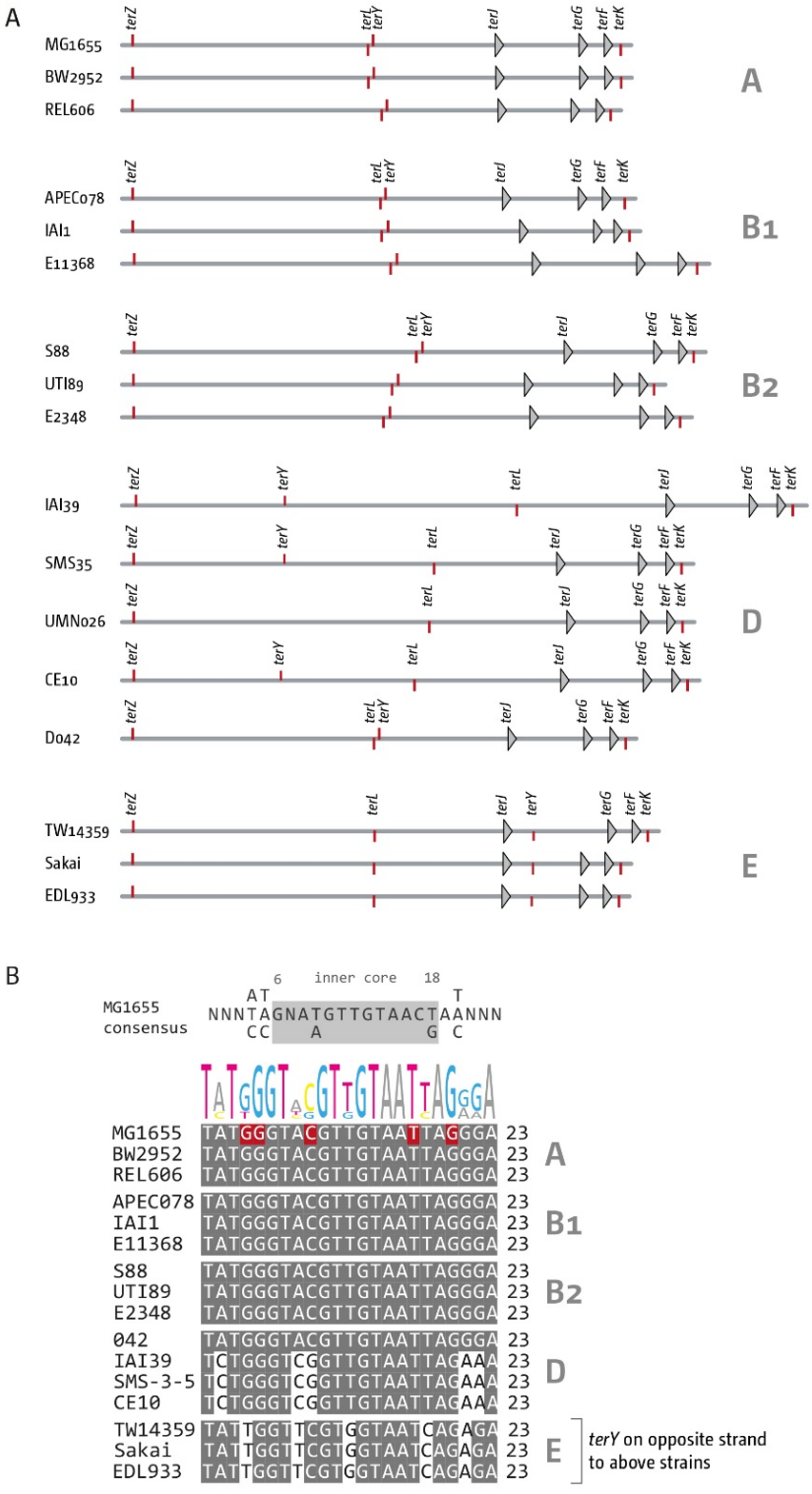

Figure 22: Distribution and sequence comparison of the secondary ter sites K, L, Y and Z in E. coli. (A) Genomic location and distribution of the secondary ter sites K, L, Y and Z across E. coli genomes clustered by phylogroup (A, B1, B2, D and E). Primary ter sites F, G and J are depicted by triangles for orientation. Secondary ter sites are shown by red tick marks, with tick marks below and above each genome line indicating permissive and restrictive orientations respectively for replication forks approaching from the left. (B) Sequence comparison of the *terY* site across the phylogenetic groups.

By analysing *terKLYZ*, known as pseudo *ter* sites (Duggin and Bell, 2009), we can observe what kind of natural variation we may expect to see in *ter* sites which are not part of the native termination area, do not bind Tus efficiently, therefore do not play a role in the primary replication fork trap. By aligning the pseudo *ter* sites with predicted coding sequences obtained from Prodigal, I was able to locate *terKLZ* and there was indeed a high level of conservation in these sites. This is not surprising as the genes differing by even one SNP would be likely to change the conformation of the gene product and therefore could result in an altered protein function. Therefore, the fact that pseudo *ter* sites *terKLZ* are within ORFs, and there is a high degree of homology, likely points to the function of genes in which *terKLZ* are situated. *terY*, on the other hand, varied substantially across the phylogroups, more so than any other *ter* site, and is not present within an open reading frame.

We could not find *terY* for UMN026 of *E. coli*, and interestingly this strain also has some variation in the outer sites, as previously mentioned. *terY* was found to be located close to *terL* in phylogroups A, B1 and B2 and did not require any mismatches to be found, when compared to the consensus sequence. *terY* was found further away from the termination area in group D, apart from UMN026 where it could not be located using the original *ter* parameters, and even in strain 042 terY is in a similar location to groups A, B1 and B2. All group D strains, excluding 042, required 5 mismatches to locate *terY*. terY was found in group E to migrated toward the termination area, now being located inside of *terJ*. This is the only group where I have found *terY*, a pseudo *ter* site not within an ORF, having moved location and now being on the opposite strand to all other strains.

These *terY* results show that the position of *terY* is not only highly variable across the phylogroups, and even within phylogroups as we see with group D, but the sequence variation is also quite high. Observing this level of variability sits well with the idea that inner sites *terABCD* may be expected to have similar level of variation if there wasn't a need to maintain their location and sequence through pressures to maintain the replication fork trap architecture.

Location of the *dif* chromosomal dimer resolution site in *Shigella*

The *dif* locus has been extensively studied and has a phenotype to which it is named. 'Deletion induced filamentation' or *dif* is primarily responsible for separation of chromosome dimers following successful replication. XerCD recombinases require homologous *dif* sites to decatenate the dimer, and cells lacking this site are unable to successfully segregate chromosomes and enable cell division. The phenotype of *xerCD* mutants is an elongated cell which never fully divides and grows into long filamentous structures. As cells cannot divide, the viable titre of such a population remains the same and the population as a whole is not viable.

Shigella have been noted as essentially being a sub-clade of *E. coli*, with their genes being included as part of the *E. coli* pan-genome and occupying the S phylogroup. *Shigella* have likely arose from many independent events from varying lineages instead of forming a compact singular groups. Therefore, we have included analysis of the replication fork trap in *Shigella flexneri*, *S. boydii*, *S. sonnei* and *S. dysenteriae*, serotypes.

The innermost *ter* sites *A-D* we matched without any difficulty, as seen in Figure 23, however, there is a higher variability in the outer *ter* sites. Interestingly, we also found that for *S. sonnei* (Ss046) all 10 *ter* sites are present, but there has been an inversion of the inner termination area placing *terA* and *terD* in the left hand replichore and *terC* and *terB* in the right hand replichore, while the other *ter* sites remain in the same position. While the order of *ter* sites has changed, the relative architecture of the termination area remains stable and balanced, and we hypothesize this arrangement is likely to still result in an active replication fork trap in Ss046. For *S. flexneri 2a* we are able to find all 10 *ter* sites, with an interesting caveat that the highly conserved C6 base pairing in *terD* has been changed to an AT pairing, suggesting an impaired Tus interaction and therefore a defect in forming the crucial mouse trap locked complex which will reduce forks becoming arrested. In *S. boydii*, we could not find *terE*, and *terF* appears to have shifted into the opposite replichore, also seen in *S. dysenteriae*. We have not analysed if these changes were based on translocation events or a change in the *terE* sequence. In addition, *terH* also appears to be missing from the fork trap in *S. dysenteriae*.

Although we see higher variability of *ter* sites in the RFT of *Shigella*, we have no indication that these *ter* sites are inactive. Experiments need to be performed to back up these hypothesis in *Shigella*. This analysis revealed an additional feature of the chromosome. In both *S. boydii*

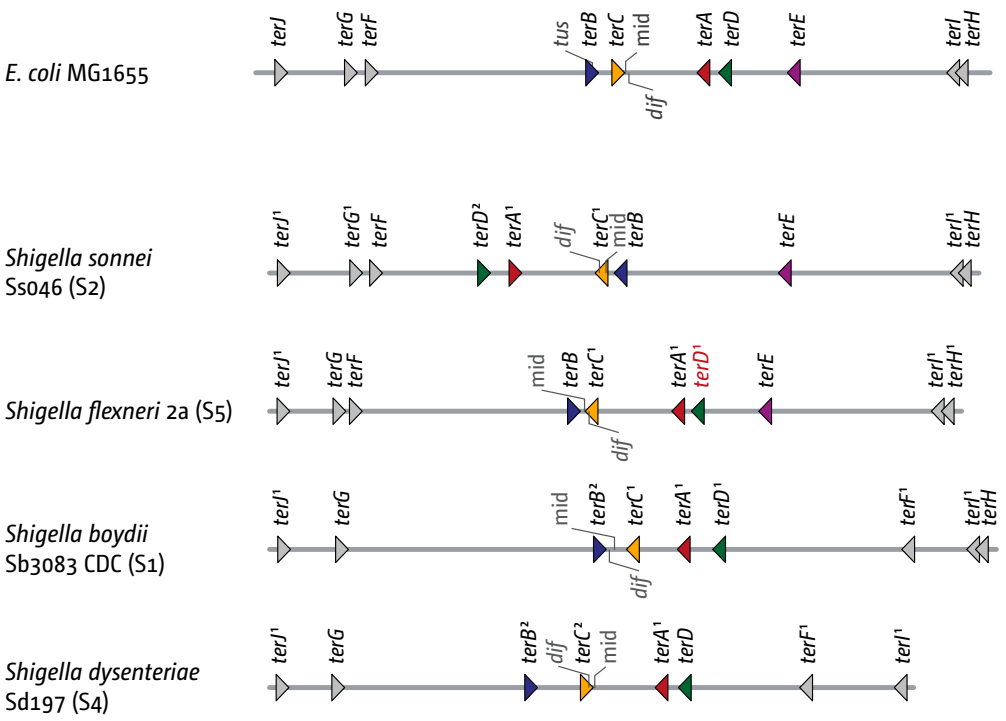

Figure 23: Replication fork trap architecture in *shigella* genomes. The fork trap architecture shows symmetry and matches with known fork trap architecture of *E. coli*, even when the region appears inverted. The position of *dif* is maintained near *terC* and the arithmetic mid point, highlighting that the relationship of *dif* to the position of *ter* sites in *shigella*.

3083 and *S. flexneri 2a* the *dif* resolution site is located between *terC* and *terB*, rather than *terA* and *terC*. In all our *E. coli* genomes *dif* is located between *terA* and *terC*, usually closer to *terC*. With these *Shigella* genomes, *terC* is now located on the opposing strand, forcing forks to fuse between *terC* and *terB*, instead of *terA* and *terC* as in *E. coli*. Therefore, even though *dif* has changed positions, it remains between the two *ter* sites where forks would frequently fuse, at the innermost *ter* sites.

It was proposed early on that forks fuse at *dif* (Hendrickson and Lawrence, 2007), based on GC skew and mutation bias analysis. However, more recent whole genome sequencing and bioinformatic work arguing against this hypothesis (Kono et al., 2011). Following this observation, we further analysed a total of 5 *S. flexneri* and 5 *S. boydii* genomes and we found one genome out of a total of five, had a RFT arrangement as seen in MG1655 (Figure 23). In all others the *dif* site is found between *terC* and *terB*, where *terC* is now in the opposite orientation causing the forks to fuse between *terC* and *terB*. Interestingly, this change often coincides with *terD* in which the CG6 base pair is substituted. Given the asymmetrical sequences of *ter* sites and *dif*, strand information is easy to identify, allowing us to draw conclusions about position of fork fusions, given the polar nature of the replication fork trap. The changes in *Shigella* architecture could have stemmed from an inversion spanning *terC* - *dif*, but we also find that for other *Shigella* genomes we see *dif* and *terC* on the same strand, a different arrangement than we see in *E. coli*. Thus, a simple inversion may not be the sole cause for this arrangement. It is tempting to speculate that *dif* location within the innermost *ter* sites is also an important feature of chromosomal architecture.

The study by Kono et al. (2011) provides further support for the importance of *dif* positioning within the replication fork trap. Their computational analysis revealed that the location of *dif* sites is highly conserved among closely related bacterial species, suggesting a crucial role in chromosome segregation and dimer resolution (Kono et al., 2011). The observed variations in the RFT architecture and *dif* location in *Shigella* genomes, compared to *E. coli* MG1655, highlight the potential significance of these architectural differences.

The *dif* site, discovered as a recA-independent recombination site in the terminus region of the *E. coli* chromosome (Kuempel et al., 1991), plays a crucial role in the resolution of chromosome dimers by site-specific recombination (Blakely et al., 1993; Sherratt et al., 1995). This process is essential for proper chromosome segregation and is tightly linked to the cell cycle (Lesterlin et al., 2004). Applying the computational approach from Kono and colleagues to analyze *dif* sites

in *Shigella* could provide additional insights into the evolutionary forces driving these changes and their impact on chromosome dynamics, particularly in the context of dimer resolution and segregation. However, this analysis was beyond the scope of the current work. Further investigation of *dif* positioning relative to the innermost *ter* sites across a wider range of bacterial genomes may shed light on the conserved and divergent features of this chromosomal architecture and its functional implications for chromosome dimer resolution, segregation, and cell cycle coordination. Such studies could also explore the potential consequences of altered *dif* locations on the efficiency and timing of chromosome dimer resolution, as well as any associated effects on genome stability and bacterial fitness.

Salmonella and *Klebsiella*

Another important question is how the replication fork trap might be structured in other bacterial species. tus-related sequences have been found in most Enterobacteriales, and it has been shown before that the expression of tus genes from *Salmonella enterica*, *Klebsiella ozaenae* and *yersinia pestis* in *E. coli* resulted in the formation of functional Tus-*ter* complexes, even though the blocking activity of the heterologous complexes were not as strong. However, the sequences of Tus from Salmonella, Klebsiella and Yersinia show substantial differences, and without a biochemical analysis it is not possible to deduce whether sequence variations seen in potential *ter* sites might lead to inactivation or are based on different interactions with the species-specific Tus protein. To get at least some insight into the structure of fork traps in other species, we investigated two Salmonella enterica genomes, as well as Klebsiella variicola and Klebsiella pneumoniae genomes. To identify potential *ter* sites, we used precisely the same parameters as for *Shigella*, allowing a maximum of five mismatches. For all genomes, we excluded *ter* sites that had a number of mismatches throughout and, in addition, lacked the critical G6. All other potential *ter* sites found were included (Figure 24 A). Given that a significant number of mismatches were observed in all putative *ter* sites (Figure 24 B), we numbered them, rather than using the nomenclature used in *E. coli*.

Work by Neylon (Neylon et al., 2005) showed bioinformatic analysis of Tus protein and aligned the MG1655 laboratory strain against other species of bacteria, including *Salmonella*, *Klebsiella* and *Yersinia*. These results showed that indeed there is variation in Tus sequence across different bacterial clades. Our analysis into *E. coli* Tus aligned showed high homology across phylogroups (Figure 24), in which those strains occupy commensal as well as pathogenic niches (Sims and Kim, 2011; Touchon et al., 2009). We finalised our bioinformatics approach to identify the architecture of the replication fork trap in *Salmonella* and *Klebsiella* based off of known Tus protein alignments.

By using Bowtie2 alignment software and Muscle alignment algorithms, we found that the replication fork trap structure looked remarkably similar to that of *E. coli*, however, as the names of each *ter* site are a matter of historical semantics we elected to omit the classical *terA-J* nomenclature for a numbering system (Figure 24). We found that there was indeed what looked to be a set of inner *ter* sites that resemble the primary *ter* sites in *E. coli* and outer sites which lay more towards the periphery. By overlaying the *dif* site and mid-point onto the chromosome maps to see and trends that we also observed with *Shigella*. Indeed, the *dif* site was retained within the

innermost *ter* sites, even though the arithmetic mid-point varied, further strengthening the notion that maintaining low copy number of *dif* is beneficial for maintaining replication efficiency (Duggin and Bell, 2009; Hill, 1992).

The exclusion criteria for *ter* sites in these two species were as follows; 1) any sites with higher than 5 mismatches needed to score a true hit and 2) any sites which lacked the crucial GC6 base pairing shown to form the so called 'molecular mouse trap' (Berghuis et al., 2018; Elshenawy et al., 2015; Mulcair et al., 2006). Sequence analysis of *Klebsiella* revealed much higher diversity compared with *Salmonella*, which is unsurprising considering *Salmonella* is a closer relative to *E. coli* (Galli et al., 2019). The *ter* sites could not be named conventionally with either species due to the diversity and amount of mismatches required to get true *ter* site matches.

Without the biochemical characterisation that the *ter* sites found are indeed active, the analysis must remain limited. However, all genomes show a pattern of putative *ter* sites that is strikingly similar to what is observed in *E. coli*. In both Salmonella genomes, we found an area opposite to *oriC* that is flanked by a tight pair of clustered *ter* sites. The chromosomal mid-point and *dif* chromosome dimer resolution site both are located between the innermost *ter* sites. We noted that the pair of *ter* sites on the left-hand side in Figure 24 is followed by another *ter* site, a pattern reminiscent of *ter* sites *A-D* and *E* in *E. coli*. Similarly, both in *K. variicola* and *K. pneumoniae* we found that the *dif* chromosome dimer resolution site was flanked by at least two *ter* sites which would form a trap. Interestingly, in these genomes, we found that the arithmetic mid-point is not located within the innermost *ter* sites. If these *ter* sites were to form functional Tus-*ter* complexes, this would be in line with our hypothesis that keeping the *dif* chromosome dimer resolution site close to the fork fusion point is one of the important functions of the replication fork trap. It will be intriguing to investigate this point further in future studies.

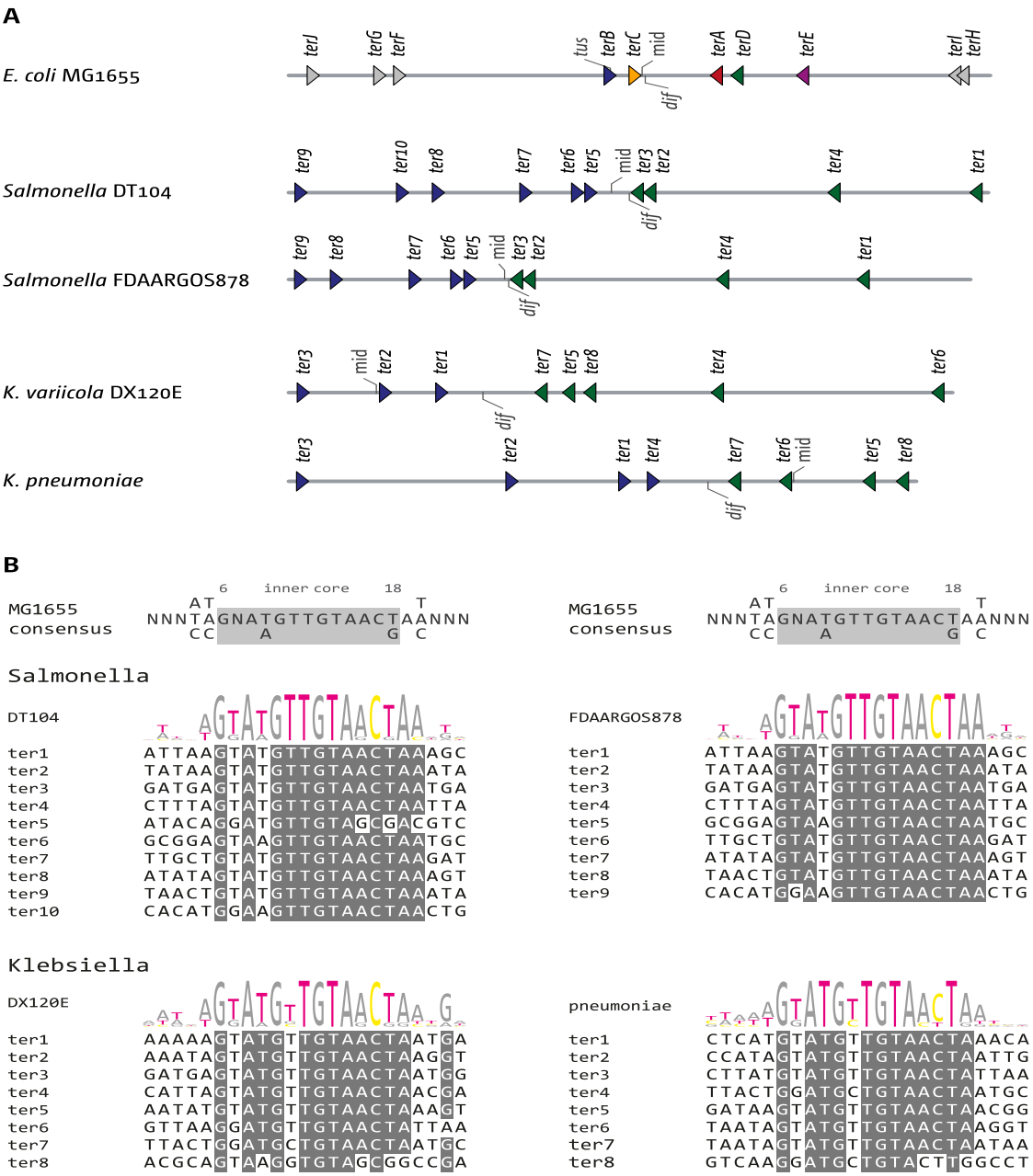

Figure 24: The fork traps were mapped in Salmonella enterica, *K. variicola*, and *K. pneumoniae* genomes, consisting of ter sites and a dif site. The ter sites were numbered and compared to an E. coli reference to assess similarities and differences. The DNA ter site sequences were compared between the Salmonella, *K. variicola*, and *K. pneumoniae* genomes, using a consensus *E. coli* sequence for base pair matching. Identical nucleotides were highlighted to visualize sequence conservation. Alignments of matched *ter* sites in *salmonella* and klebsiella genomes. A) The fork trap architecture. B) The *ter* site alignments using the MSA package in R, from Bioconductor.

Chapter 1 - Discussion

The results show that surprisingly most *ter* sites show high degree of conservation across the species. Following gene prediction using Prodigal, the inner *ter* sites *ABCDE* were shown not be a part of open reading frames (ORFs) whereas all others were part of ORFs. Gene sequence conservation obviously requires high fidelity to maintain the function of the gene product so therefore it is unsurprising that those sites within ORFs are highly conserved. Work from Toft and colleagues (Toft et al., 2021) showed the outer sites are rarely bound by Tus using Chip-Seq, confirming as Duggin and Bell did (Duggin and Bell, 2009), that *terFGHIJ* are indeed pseudo sites and not a part of the functional fork trap in *E. coli*. Surprisingly, *terABCDE* were highly conserved throughout the analysed genomes even though they are not within ORFs.

This chapter addresses the fact that all previous RFT studies have been performed in K12, the first sequence being MG1655, these data illustrate a more conserved fork trap architecture across the Enterobacteriaceae than was initially hypothesized. The location of outer *ter* sites within ORFs and the inner sites not within ORFs strengthens the idea that the conservation of these outer sites are mainly down to the gene in which they reside. Inner sites were expected to have greater variation in their sequence and orientation depending on the strain/species, however there was remarkable conservation across the entire phylogenetic group, showing the importance for maintaining the fork trap once it has been acquired.

The data presented in this chapter show the importance of maintaining the sequences and architecture of the replication fork trap across phylogenetic groups of *E. coli*. Recent work by Zhang and colleagues (Zhang et al., 2023) has shown that chromosome organization shapes replisome dynamics in *Caulobacter crescentus*, highlighting how the unique fork trap system in bacteria likely influences replication fork coordination.

The fully conserved inner fork trap sites across *E. coli* groups indicate this system is universally present, not just in the MG1655 laboratory strain. Bioinformatic analysis also suggests the fork trap system is active in closely related species like *Shigella* and *Salmonella* based on high Tus protein sequence similarity. Considering the strains tested here range from commensal laboratory strains to the infamous O157:H7 enterohaemorrhagic strain, it is surprising that the fork trap is maintained so strongly across these different niches, at least showing that the fork trap being present is not down to the environment in which the bacteria occupies, but the fork trap address a universal need for termination control. In *Shigella*, the positioning of the *dif* site next

to the innermost *ter* site further demonstrates the importance of co-localizing the dimer resolution site with the fork fusion point. Keeping *dif* at a low copy number is crucial for proper chromosome segregation, and any over-replication of the *ter* region would increase *dif* copies and create defects.

The mild growth defect of *Δtus* mutants indicates the fork trap is not essential, but consistently present across species analyzed. The autoregulated *tus* gene is a defining factor of the system. Recent work by Toft et al. (Toft et al., 2021) provides evidence against the model of outer *ter* sites acting as 'backup' sites, and instead these outer sites are likely conserved purely due to being located within ORFs. Pseudo-*ter* sites like *terY* are more variable in terms of location and sequence, indicating a lack of selection pressure to maintain sites not participating in the primary fork trap. This analysis supports the hypothesis that non-essential *ter* sites are susceptible to sequence variation and displacement across the chromosome.

The biochemical interactions between Tus and genomic *ter* sites have been studied extensively, with the precise amino acids required for successful replication fork blocking coming from a number of notable works (Bastia et al., 2008; Berghuis et al., 2015, 2018; Elshenawy et al., 2015; Neylon et al., 2005). Despite this, the precise function of the replication fork trap remains elusive. A replication fork trap system is present in most Enterobacteriales, pseudoalteromonas and aeromonales, with the Bacillales also showing a fork trap system lacking protein sequence homology to Tus, indicating that this system has likely evolved via convergent evolution, suggesting that the presence of a fork trap has an important role in DNA replication.

Cells lacking Tus show a very mild growth defect at best and the trend was shown to be significant across seven data sets, re-writing the view that *tus* mutants lack a noticable phenotype (Hyrien, 2000). The majority of fusions take place at the midpoint, away from Tus-*ter*, with early hypothesis being that the RFT prevents forks from entering the opposing replichore to prevent head on replication-transcription collisions, but this does not require a large RFT spanning ~45% of the chromosome, and B.subtilis is evidence of this with ~5% coverage of *ter* sites. So, it is likely that the now recognized pseudo *ter* sites are conserved based on their presence in ORFs. This idea has been strengthened recently with the experiments by Toft and colleagues who argue against the idea of back up *ter* sites being necessary for a functional fork trap system ((Toft et al., 2021)) and call for all *ter* sites except *terA-D* to be redefined as pseudo sites. Indeed Duggin and colleagues (Duggin and Bell, 2009) do show that the outer *ter* sites (*terF-J*) are bound by Tus protein, albeit with reduced binding efficiency as shown by single molecule anal-

ysis using hairpin DNA structures and magnetic bead experiments (Berghuis et al., 2015), but there is little evidence that these outer *ter* sites participate in the stalling of forks attempting to escape the termination area and so may be thought of as being pseudo *ter* sites being conserved across evolution purely from their location being within open reading frames.

It would be interesting to compare the genes in which the outer *ter* sites sit in *E. coli* with those of *B. subtilis* to see any differences can be found and if remnant *ter* sites could be identified in these genes. As pseudo-*ter* sites are highly conserved due to the presence in an ORF, however *terY* is more variable in terms of location, nucleotide substitutions and strand, therefore blocking direction.

Shigella have a more variable RFT as they have evolved independently through a number of lineages. The inner termination area shares high similarity with *E. coli*, despite this fact. Forks that are stalled at Tus-*ter* in strains with excess Tus require *recA* filament recombinase to bypass the barrier and proceed with replication, thus viability. GCR can be observed as suppressor mutations that flip the *ter* sites to become permissible in some strains to allow the movement of a fork, showing how substantial Tus is to the replisome. If there was no pressure to keep *ter* sites then the chromosome would have lost them or increased variability, as seen with *terY*. This analysis is in line with my hypothesis that *ter* sites lacking strong Tus affinity that do not participate in the primary fork trap, therefore without a strong selection pressure, will be susceptible to natural sequence variation and prone to displacement across the chromosome where other chromosome dynamic pressures out-weight the need to maintain the location of the *ter* site. It is therefore likely that terY is a relic *ter* site which is no longer required to maintain chromosome stability and throughout evolution of the species there were little consequences of this site changing over time.

Some changes observed in specific strains, such as an apparent swapping of *terAD* or *CB* in SMS3-5 and IAI39, still maintains the inner fork trap system, likely with no consequence. Highlighting the importance of fork trap symmetry in maintaining functionality. Inactivation of Tus-*ter* in plasmid R1 showed aberrant structures such as rolling circle replication (Krabbe et al., 1997) and that loss of Tus-*ter* showed segregation defects and plasmid loss, thus a selection pressure, and even Hamilton and colleagues recently showed that the incorporation of *ter* sites into their plasmid system mitigated aberrant recombination in which plasmid instability was dependent (Hamilton et al., 2023).

Galli and colleagues (Galli et al., 2019) showed that in Pseudoalteromonadales the *tus* gene is

associated with chromosome II, of plasmid origin. The Tus gene is within the *terB* site, acting as auto-regulator through negative feedback and acts in the same way across all species with a known fork trap. Gene autoregulation is a key feature of fork trap architectures, which seems to be convergent feature as well seeing as this is also the case for *B. subtilis* and it's fork trap.

My findings are consistent with other findings from Rudolph and collegues, that intermediates formed as a natural part of termination can trigger aberrant replication that must be delt with, the specifics of which are tackled in my other chapters. The idea here is that the replication fork trap contains the over-replication to prevent it from bleeding out of the termination area and back into the chromosome. This would not only increase replication-transcription collision but also increasing the copy number of key areas along the chromosome, altering the balance of transcribed genes. Therefore there may be a role for the replication fork trap in regulating gene expression and could help explain our analysis showing that RNA metabolic genes are excluded from the termination area.

The model proposed in (Midgley-Smith et al., 2019), and other works, suggests that one of the main intermediates that is responsible for the this pathological over-replication is a 3' flap which can be the target of PriA restart protein. This structure would be degraded by 3' exonucleases such as ExoI, ExoVII and SbcCD, as well as converted from a 3' flap into a 5' flap by RecG helicase, the subject of which is discussed further in later chapters. This intermediate would form if the DnaB replicative helicase displaces the leading strand as it translocates along ssDNA, and would then need to be removed.

The location of crucial genes, such as those involved in RNA metabolism, would therefore be subject to nuclease activity if they were to reside within the termination area which could result in catastrophic consequences and pose a risk to cell viability. There could also be a gene dose effect from over-replication within this area, as already discussed. It has been well established that *Shigella* are essentially *E. coli* which evolved from a different genetic branch. My research into *Shigella* genomes revealed an interesting fact that could add an interesting angle to the story. The location of the *dif* dimer resolution site was shown to correlate with the location of the inner *ter* sites. XerCD recombinase requires two *dif* sites to decatenate the chromosome upon successful duplication to facilitate the final stages of binary fission. Naturally, XerCD will require the lowest possible copy number of *dif* in order to resolve chromosomal dimers correctly, should there be more copies of *dif* then a connection between sister chromosomes may remain at an ectopic site even after XerCD have resolved the dimer.

I have used a strain with an ectopic *dif* site to test the doubling time when there are more copies of *dif* than are needed, and indeed the strain has an extended doubling time. Keeping *dif* to a low copy number is therefore of paramount importance to resolution of chromosomal dimers following successful replication. Even in bacteria with two chromosome, such as *vibrio Cholerae*, the two *dif* sites on each chromosome vary in their *dif* sequence to reduce the liklihood of recombination between the two chromosomes. Previous bioinformatic work, as well as my own, show that the *dif* site is not the same as the arithmetic mid point along the chromosome. The *dif* site does appear to have a strand preference and sits next to *terC* in the majority of strains analyzed, which was observed in *Shigella* genomes where the termination area is inverted. Any over-replication of the termination area which increases the copy number of *dif* would therefore need to be removed rapidly should the cell be able to segregate chromosomal dimers.

The data provided in this chapter show that the replication fork trap has evolved in likelihood to contain unwanted intermediates formed during the process of fork fusions. Due to the momentum and speed of each replication forks, it would be logical to assume that this replication fork 'crash' results in a type of DNA debris which can be site for PriA mediated replisome loading and aberrant replication. These data therefore dismiss the notion that the replication fork trap has evolved to simply direct the fusion of forks into a specified area. It is likely that the fork trap has evolved around the natural mid-point of the chromosome and fork fusion location, rather than the other way around. This trap prevents forks originating from *oriC*, or anywhere for that matter, from progressing back into the chromosome where they will encounter increased replication-transcription collisions which can be the source of genomic instability through stabilization of R-loops, discussed in chapter 3.

Although maintaining codirectionally of replication and transcription would be a welcomed benefit to maintain normal chromosome dynamics, it seems this is not the sole function of the replication fork trap and is unlikely to be a strong driver in Tus-*ter* evolution, shown primarily by the lack of a strong growth phenotype in Tus single mutants. Recent work by (Jameson et al., 2021) and colleagues also show evidence for under replication when forks are forced to fuse at *ter* sites when one fork is stalled for some time before the other meets it, highlighting the fact that our picture of termination and the replication fork trap system is still incomplete.

There are likely other factors at play, such as the positioning of the *ter* domain at mid cell during the final stages of chromosome segregation. It is know that MatP binds to the chromosomal *ter* domain and interacts with the divisome, but also inhibits MukBEF activity, which forms struc-

tural loops within the chromosome and MatP inhibiting this chromosome structure would also assist locating the *ter* domain at mid-cell (Mäkelä and Sherratt, 2020). It appears there is no one reason for the evolution and presence of the replication fork trap in *E. coli* and other species as shown, but rather that the fork trap system confers multiple interconnected benefits to reduce chromosme instability that allows for proper dimer resolution and chromosome segregation.

Our work has shown that the RFT is conserved across the *E. coli* phylogenetic groups as well as *Shigella*, Salmonella and Klebsiella (Goodall et al., 2021). Using bioinformatic approaches to analyse the presence of primary and secondary *ter* sites, we have concluded that a fork trap is likely active in all genomes analysed (17 *E. coli*, 11 *Shigella* and 2 *Salmonella*, 2 *Klebsiella*). Given the relatively large replication fork trap seen in *E. coli*, which spans rouhgly 45% of the chromosome, and the comparatively narrow fork trap area in *B. subtilis* which spans roughly 9% of the chromosome, we have also strengthened the idea that the inner *ter* sites of *E. coli* constitute the primary fork trap area and the outer *ter* sites can be thought of as secondary, given their relatively low binding affinities for Tus protein, and also that pseudoTer sites ($terK, L, Y, Z$) are likely not used as the main fork trap in *E. coli* (Duggin and Bell, 2009)

This research has shown that the fork trap is conserved among all phylogroups of *E. coli* which includes both commensal, pathogenic strains and in *Shigella*, strengthening the idea that this system provides some benefit to cells, and once obtained in the bacterial genome it is maintained. With the newly generated growth curve data in *tus* cells compared with MG1655, I have also shown that there is a consistent growth delay in the *tus* single mutant over all data sets, although this trend is slight between 0.5 - 1 minutes delay difference. This coupled with the mild over-replication in *tus* single mutants (Rudolph et al., 2013) indicates that there could be more going on in cells lacking a replication fork trap than was previously thought, further research probing the consequences of fork fusions that might result in genomic instability will all contribute to uncovering more important details about the purpose of the RFT in *E. coli* and the process leading to successful DNA replication termination.

Future Work

Greater expansion of fork trap architectures

Our results reveal two distantly related bacterial species, *E. coli* and *B. subtilis*, although the latter hasn't been investigated here, that utilize a fork trap system to regulate DNA replication

termination, utilizing Tus and RTP proteins, respectively. While the fork trap is a conserved mechanism, the components involved can vary between bacterial clades. Further phylogenetic analyses probing deeper branches of the bacterial domain could uncover additional fork trap systems that diverge from the canonical Tus-*ter* model. Identifying other fork trap components and characterizing their molecular mechanisms would provide greater insight into how the fork trap arose and diversified over evolutionary timescales.

In particular, investigating whether other bacterial species possess a distinct but analogous fork trap complex could elucidate broader principles governing the interplay between chromosome structure and replication regulation. The organization of the termination regions differs between Tus- and RTP-utilizing species, suggesting fork traps adapted to distinct chromosomal architectures. Uncovering a greater diversity of fork trap systems may reveal generalizable patterns in how termination sites are specified and integrated with global chromosome structure and topology.

Overall, expanding the phylogenetic tree of fork trap systems will advance our understanding of both the purpose these conserved mechanisms serve and the ability of bacteria to modulate a universal cellular process to suit their specific genomic contexts. While Tus and RTP represent two distantly related implementations, probing deeper into bacterial diversity could uncover other variants of the fork trap at key evolutionary branch points. Characterizing these systems at the molecular level would clarify the essential aspects of the fork trap that remain conserved versus the aspects that exhibit flexibility to adapt to different bacterial lifestyles and chromosomal designs.

Chapter 2: Consequences of Fork Fusions and Recombination Rates

The aims of this chapter are to measure revertant frequencies and calculate recombination rates based off of fluctuation data from many cultures grown in parallel. To determine the role, if any, of the replication fork trap in recombination rates inside the termination area. To assess the impact of RecG and UvrD helicase on recombination rates at fusion sites and finally to provide mechanistic insight into how the reverted cells are generated by measuring cells lacking RecA recombinase.

Faithful completion of DNA replication each cell cycle is essential for maintaining genomic integrity in all domains of life. Replication initiates bi-directionally from a single origin (*oriC*) and concludes when the two replication forks converge at the opposing terminus region. This fork fusion process is inherently risky, as any problems arising during elongation leave little margin for fork restart or repair. The *E. coli* terminus contains binding sites (*ter*) for the Tus protein, which forms complexes that trap replication forks to funnel termination events into a chromosome specific location. The evolutionary drivers behind this architectural system have been suggested as prevent over-replication of the termination area, keeping the *dif* dimer resolution site to as low a copy number as possible for accurate decatenation, and minimizing replication-transcription conflicts that can occur if forks were to move out of the termination, something observed when *tus* is deleted inactivating the fork trap. However, regulated fork pausing by Tus-*ter* can cause issues if the two forks become stalled asynchronously, which readily occurs when one fork encounters a barrier while the other continues unimpeded (Duggin et al., 2008). Fork trap systems were proposed to induce homologous recombination when forks become oppositely arrested, likely as a mechanism to restart or repair stalled replication (Duggin and Bell, 2009) and early termination research coined Tus-*ter* barriers as a 'hot spots' of recombination (Horiuchi et al., 1994).

Recent *in vitro* experiments show a reduction in level of overall recombination when *ter* sites were placed at locations where forks are fusing (Hamilton et al., 2023). Therefore, do Tus-*ter* complexes and the replication fork trap induce or mitigate recombination inside the termination area? The termination area has been implicated in causing recombination through Tus-*ter* fork pausing barriers, but also Hamilton and colleagues showed recombination is lowered on their plasmid model when *ter* sites and a fork trap is present. Based off of these two conflicting ideas, my research aims to provide a clearer picture of recombination inside the termination area, as well as other sites in the *E. coli* chromosome (Midgley-Smith et al., 2018).

Fork fusion events are known to stimulate pathological over-replication and these effects are exacerbated in combinations of *recG* and *3' exonuclease* mutants, as well as *rnhA* and *tus*, showing the complex array of proteins which are at play inside the termination area to process aberrant structures that arise from replisome induced strand displacements during fork fusion (Dimude et al., 2016; Midgley-Smith et al., 2019).

The *E. coli* terminus is a recombination hotspot, but the drivers of this localized hyper-recombination are unclear (Horiuchi et al., 1994). One hypothesis is that Tus-*ter* complexes directly stimulate recombination upon fork arrest. However, unregulated fork collisions could also generate problematic joint molecule structures like 3' flaps or Holliday junctions that require processing by recombinases (Michel et al., 2018; Rudolph et al., 2013).

I test the hypothesis that fork fusions themselves drive increased recombination rates by looking at *tldD* and *yjhR* locations in the chromosome in double origin strains. We are able to infer the direct impact of fork fusion events as *tldD* and *yjhR* locations do not contain fork trap architecture. This would best mimic the scenario we would observe in a bacteria without a fork trap.

The restricted architecture of bacterial chromosomes, with a single origin and terminus, necessitates intricate coordination of replication completion. By trapping forks, Tus-*ter* prevents collisions in distal regions that could jeopardize genomic integrity. However, regulated fork pausing poses risks if replication becomes asynchronous. Multiple accessory helicases and nucleases help resolve joint molecules generated when forks meet at Tus-*ter* or fuse freely (Lloyd and Rudolph, 2016; Wendel et al., 2014) and I test two helicases of interest, RecG and UvrD, in this scenario at different chromosomal locations. Future work should explore whether fusion at Tus-*ter* elicits distinct intermediates or responses compared to unrestricted collisions. By understanding the effect a RFT system has on rate of recombination, we can formulate an idea

about how this system may have evolved to preserve chromosomal stability and reduce events which inherently reduce genome duplication fidelity.

Outline of experiments and description of the cassette

These experiments are designed to test the rate of homologous recombination within a strain by incorporating the *kankanMX4* tandem repeat cassette into one location on the chromosome. The *kankanMX4* cassette contains an internal 266bp duplication of the resistance gene for kanamycin with a stop codon in the middle of the open reading frame (Figure 25). The kanamycin resistance is restored if the chromosome loses one the tandem repeats, therefore correcting the truncation and restoring resistance to kanamycin. The cassette is placed and one of three sites, *narU*, *yjhR*, and *tldD*, which lay inside the native termination area, in the right hand replichore, or the left hand replichore, respectively.

When one reverted cell arises, termed a revertant, every progenitor cell will have full resistance and will grow in the presence of kanamycin. Due to the random nature of these events, those which happen earlier during growth will naturally result in a higher number of progeny with kanamycin resistance and those that occur later will have fewer resistant progeny cells inside the culture. As we cannot know when a particular recombination event takes place within a culture, we performed fluctuation analysis as stated in Foster (2006) which builds on from Lea and Coulson's median culture calculations (Lea and Coulson, 1949) but optimised and modernised by using the Flan package in R and builds upon the methods used in the experiments by Swings and colleagues (Swings et al., 2017).

oriX and *oriZ* are both additional ectopic origins which are placed roughly halfway into the left and right replichores, respectively. In an $oriC^{+}$ $oriZ^{+}$ background the fusion point is proximal to the *yjhR* gene in *E. coli*, whereas in a $oriC^{+}$ $oriX^{+}$ background the novel fusion point is now proximal to the *tldD* gene. Integrating the *kankanMX4* cassette in these locations are based on the replication profiling and using marker frequency analysis from whole genome sequencing, sensitive for detecting low copy regions known to be fusion points in WT, $oriZ^{+}$ and $oriX^{+}$.

Three strains constructed from the pSLM001 plasmid, which carries the *kankanMX4* cassette (Sarah Midgley-Smith, 2018). One with the cassette placed near *yjhR* proximal to *oriC* in the right hand replichore and another placed in the left hand replichore near *tldD*, finally one with the cassette placed near *narU* proximal to the termination area (Figure 26).

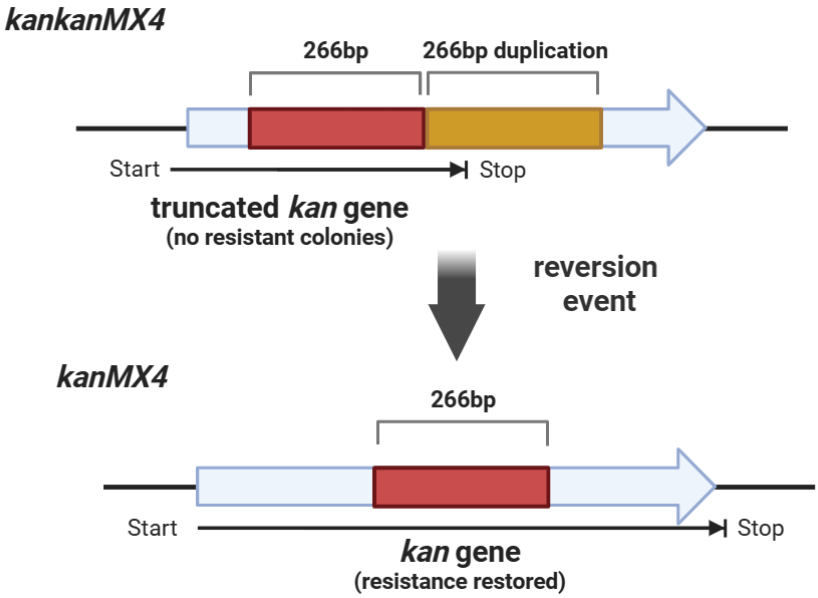

Figure 25: Illustration of the *kankanMX4* cassette with an internal tandem repeat of 266bp causing premature truncation of the kanamycin resistance gene product. Image was rendered using BioRender.

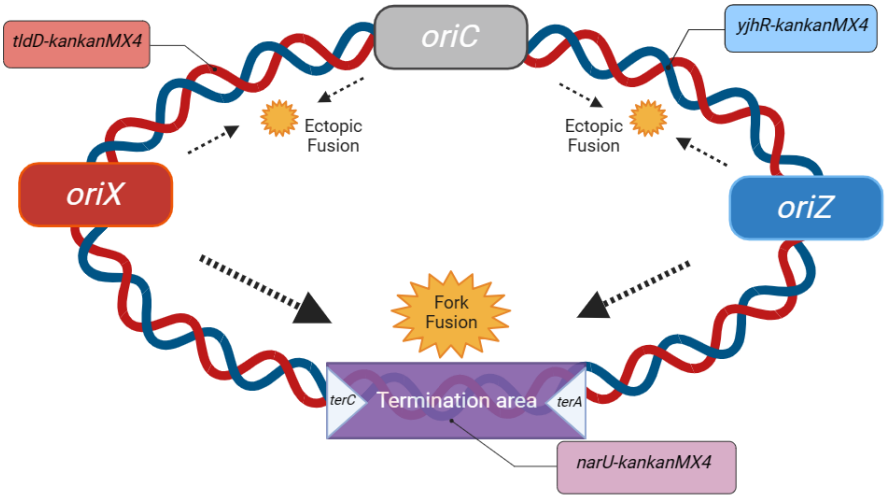

Figure 26: Chromosome positions of cassettes, origins and fusion points shown. Strains used in the experiment were WT, *oriC*$^{+}$ *oriX*$^{+}$ and *oriC*$^{+}$ *oriZ*$^{+}$. Strains contained only one copy of the *kankanMX4* cassette, at one of the defined locations. The schematic shows all possibilities of the chromosome arrangement for simplicity. Image rendered using BioRender,

Fork fusions increase recombination rates at engineered ectopic fork fusion sites

Recombination rates at ectopic fork fusion sites in the right hand replichore (*yjhR*)

In *E. coli*, the fork trap restricts the location in which replication forks can fuse in the final stages of DNA replication. I wanted to test the hypothesis that forks fusions themselves are inherently unstable that perturb local genomic instability. To test consequences of fork fusions, I wanted to measure recombination rates in an area of the chromosome which has not evolved any site specific mechanisms to deal with free fusing replications forks, in order to reduce noise that these mechanisms might have on forks fusing near *narU* in the termination area. In *oriC*$^{+}$ *oriZ*$^{+}$, this location is near the *yjhR* gene.

Measurement at *yjhR* in *oriZ*$^{+}$ strains showed a significant increase in recombination rates, 2.8 fold, compared with the single origin *yjhR-kankanMX4* (Figure 27). If this effect is dependent on fork fusions, then measurement at *yjhR* in an *oriX*$^{+}$ background should not have the same increase. Indeed, the recombination rate increase was not observed in the *oriX*$^{+}$ *yjhR-kankanMX4* construct, showing that it is the activity of fork fusions themselves that are increasing the rate at *yjhR*.

Strengthened from the *oriX*$^{+}$ data where we still measure the rate of recombination at *yjhR*, and shows that the level is comparable single origin strains, illustrating that the presence of additional origins in the chromosome does not increase the reversion rates in general.

The average of the rates calculated from four independent experiments showed that recombination at *yjhR* in an *oriZ*$^{+}$ background where the ectopic fusion point is located was increase 2.8 fold. The rates at *yjhR* in an *oriX*$^{+}$ background, on the other hand, show a more similar frequency and rate to WT, demonstrating that fork fusions increase incidence of *kankanMX4* reversion in this area of the chromosome. These data support the hypothesis that fork fusions are leading to increased rates of reversion at the *kankanMX4* cassette.

The addition of an ectopic origin into the chromosome, *oriZ*$^{+}$, sets up an additional ectopic termination area between *oriC* and *oriZ*, near the *yjhR* gene locus. In the *oriZ*$^{+}$ *yjhR-kankanMX4* construct, forks are actively fusing near *yjhR* so measuring recombination rates in this background gives us insight into the effect of freely fusing forks that do not have any site specific

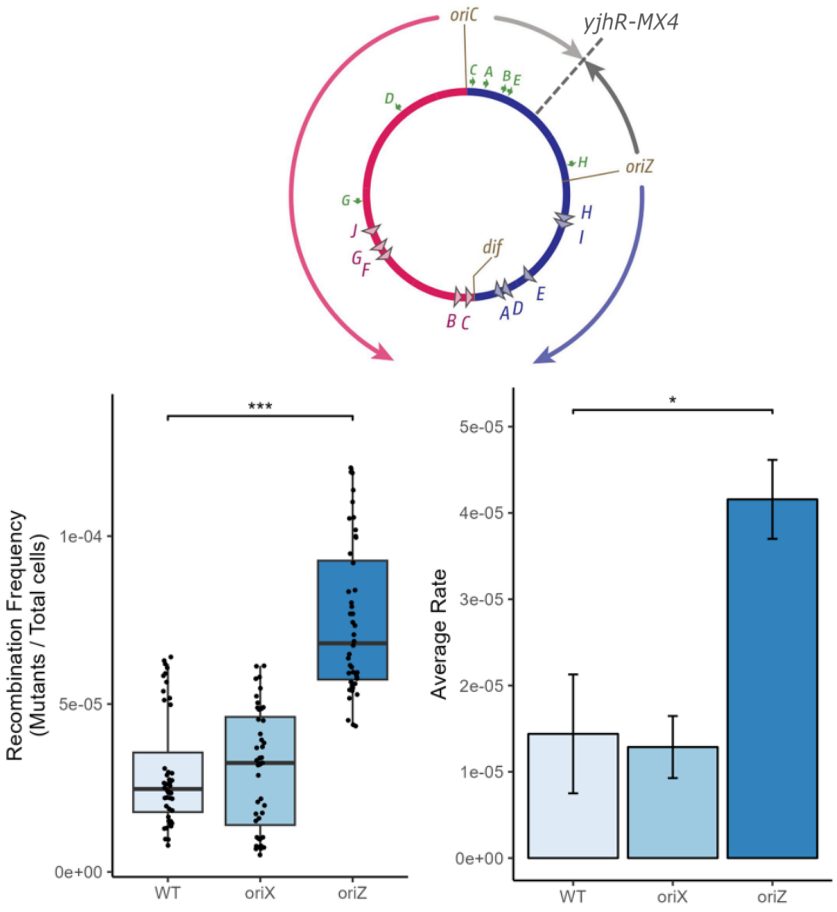

Figure 27: Frequency and rates measured at *yjhR* in the right replichore, where forks fuse in *oriZ* background, showing increased reversion frequencies and rates only in *oriZ*$^{+}$ arrangement. Figure showing 44 individual data points for each strain as frequencies, across 4 independent experiments, where rates were calculated at the end of each experiment. Rate bargraphs show average of the 4 rates and error bars showing SEM and ANOVA test was performed on each group of data points to determine p values.

means to curb consequences of fork fusions.

Recombination at ectopic fork fusion site in the left hand replichore (*tldD*)

tldD-kankanMX4 strain needed to be constructed from the ground up using recombineering to insert the cassette next to the *tldD* gene, roughly at the midpoint between *oriC* and *oriX*. *oriC*$^{+}$ *oriX*$^{+}$ strains recombineered for *tldD-kankanMX4* integration to test the hypothesis that it is the fusion of forks contributing to this significant increase, and not something down to the local area in the right hand replichore.

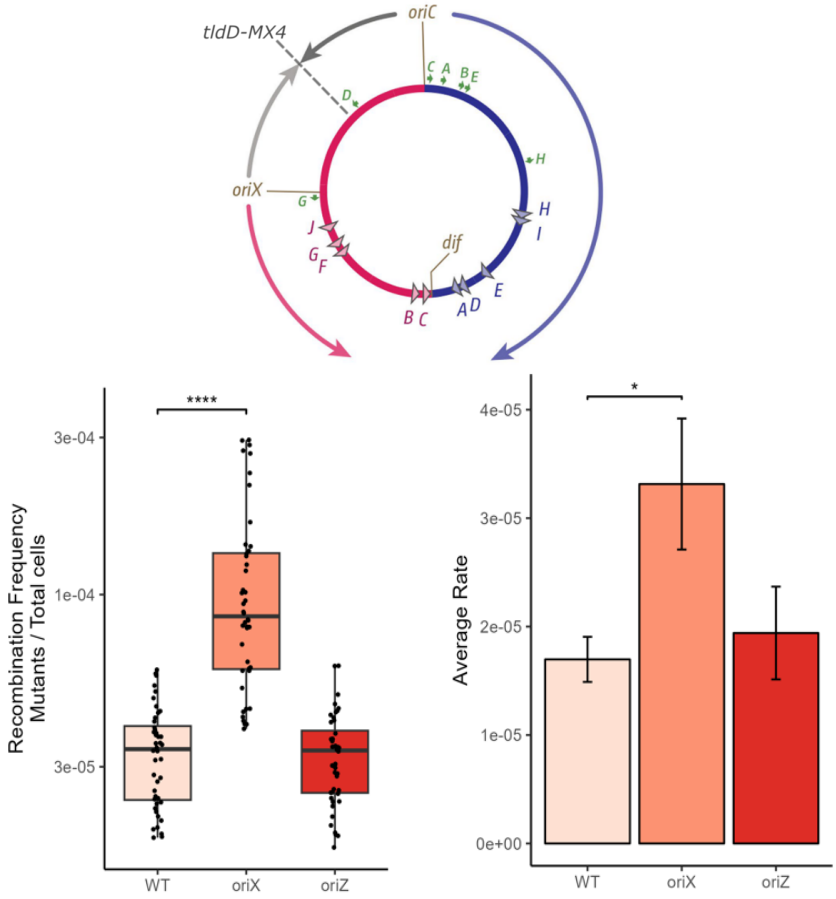

Figure 28: Frequency and rates measured at *tldD* in the left replichore, where forks fuse in *oriX* background, showing increased reversion frequencies and rates only in *oriX*$^{+}$ arrangement. Figure showing 44 individual data points for each strain as frequencies, across 4 independent experiments, where rates were calculated at the end of each experiment. Rate bargraphs show average of the 4 rates and error bars showing SEM and ANOVA test was performed on each group of data points to determine p values.

This analysis shows that *oriX*$^{+}$ *tldD* strains have the highest distribution and mean values of mutant cells and rates of recombination (Figure 28). Frequencies were statistically significant comparing both *oriX*$^{+}$/*Z*$^{+}$ to WT, where the hypothesized increase at *tldD* in *oriX*$^{+}$ backgrounds were observed. A comparison between *oriZ*$^{+}$ *yjhR* and *oriX*$^{+}$ *tldD* can therefore be made (Fig-

ure 28). The rates from four independent experiments were averaged and error bars calculated from the standard deviation of the mean. *oriX*$^{+}$ *tldD-kankanMX4* showed the highest rates across all three experiments, averaging out to a 2.5 fold increase in recombination rates when compared to the WT. *oriZ*$^{+}$ *tldD-kankanMX4* averaged out to a slightly higher rate, although this was not a significant increase.

From the averages of all the rate calculations, recombination was observed to be increased by 2.8 fold and 2.5 fold in the right and left hand replichores, respectively.

Impact of the fork trap on recombination

As the Tus-*ter* fork trap system sets the fork fusion boundaries, I wanted to measure the rate of cassette reversion in the native termination area in WT and *tus*$^{-}$ cells.

Inactivating the RFT had modest effects on rates of recombination of single origin strains inside the termination area, through measurement of kanamycin resistance in *narU-kankanMX4*. Double origin Δ*tus* strains showed reduced frequency of revertants, which would be expected if the level of recombination is heavily influenced by fork fusion events, as this background will cause forks to fuse well into the left hand replichore due to the fork trap being inactive. This shows that the replication fork trap has an effect of curbing recombination.

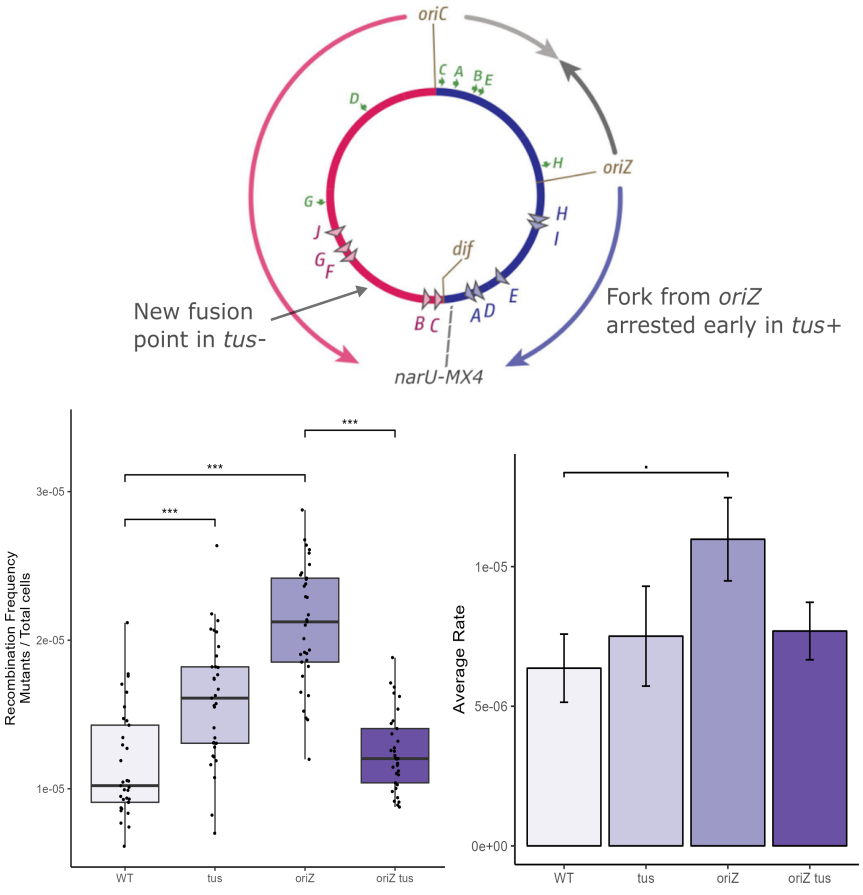

Figure 29: Frequency and rates measured in the termination area, where forks fuse in WT cells and *oriZ*$^{+}$ cells have a fork prematurely arrested at *terC*. In the *tus*- strain fork fusions are now repositioned away from *narU* into the left hand replichore. Figure showing 44 individual data points for each strain as frequencies, across 4 independent experiments, where rates were calculated at the end of each experiment. Rate bargraphs show average of the 4 rates and error bars showing SEM and ANOVA test was performed on each group of data points to determine p values.

In an *oriC*$^{+}$ *oriZ*$^{+}$ Δ*tus narU-kankanMX4* background, forks will move through the termination

area coming from *oriZ* unhindered until they meet the opposite replication fork coming from *oriC*. As determined by MFA, this fusion point is now away from the normal fusion point and well into the left hand replichore.

In *oriZ*$^{+}$ *tus*$^{-}$ background the fusion point is now well into the left hand replichore, illustrating that when forks are fusing elsewhere, there is a reduction in the frequency of a revertant developing, and also the rate drops (Figure 29). It is clearer to see a difference in *tus*$^{-}$ mutants when interpreting the frequency data, rather than the rate data, showed a more significant increase ($p = 0.00002$). This shows that there may be an increase in the rate of recombination in *tus*$^{-}$ mutants, but the rate calculation did not detect a significant enough of a difference to say with high confidence that the absence of Tus increases the rate of recombination (t-test rate p value = 0.6), only that the frequency of generating a revertants is significantly increased (Figure 29).

Measurement of recombination rates at the native *narU* locus revealed unexpected trends relating to the replication fork trap (RFT). Compared to wildtype, inactivation of the RFT led to increased recombination at *narU*, contrary to the notion that forks held up at Tus-*ter* barriers lead to increased recombination (Horiuchi et al., 1994; Sinha et al., 2018). We also observed elevated recombination rates at *narU* in *oriZ*$^{+}$ and *oriX*$^{+}$ strains with an active RFT, in which this model would come into place with forks becoming stably arrested at *ter* sites.

The role of RecG and UvrD in fork fusion induced reversion

It has been shown that over-replication occurs in multiple chromosome location in cells lacking RecG helicase, with the proposed model being the generation of 3' flaps that act as restart substrate (Rudolph et al., 2010c, 2013). The working model puts RecG as a major player, so I first asked the question of what role RecG has in fork fusion-dependent increase of cassette reversion at *yjhR*, then later in the termination area. Also, considering UvrD's role in associating with the replisome to help overcome fork blockage, alongside the evidence for UvrD in dismantling RecA from 3' ssDNA, I wanted to also see the level of revertant in *uvrD*$^{-}$ mutants.

There are overlapping criteria that make these two helicases good candidates for using to measure consequences of fork fusions. Given both are helicases with evidence for operating inside the termination area, both are essential for viability in *rnhA* mutants (Dimude et al., 2015; Hong et al., 1995), associate with SSB, and are both known to be hyper-recombinogenic (Arthur and Lloyd, 1980; Florés et al., 2005; Veaute et al., 2005), I wanted to also see the effect of *recG*$^{-}$ *uvrD*$^{-}$ double mutants on reversion of the cassette at *narU*.

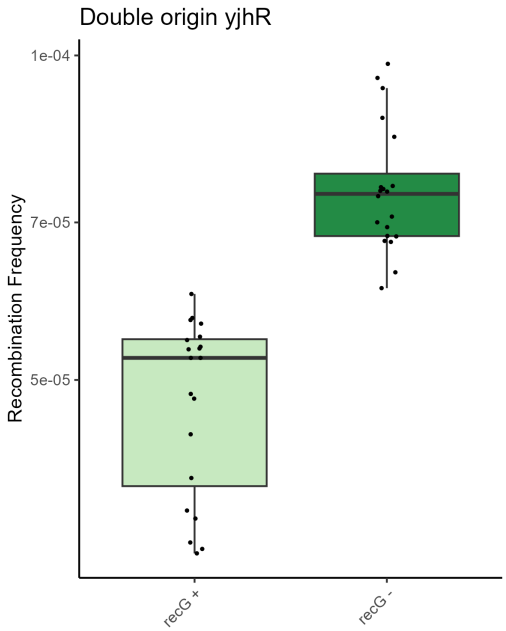

Figure 30: Frequency measured at *yjhR* in a double origin background lacking RecG. With fork fusions taking place in this area, the levels are significantly higher illustrating the role of RecG importance for homologous recombination and maintaining stability at a locus where forks are actively fusing.

In the context of replication fork fusion events at the *yjhR* locus, the absence of RecG in *oriZ*$^{+}$ strains led to a significant increase in recombination frequencies and revertant formation, as evidenced by the preliminary data from *recG*$^{-}$ mutants (Figure 30). This observation high-

lights the critical role of RecG in modulating recombination outcomes during fork fusion. RecG, a highly conserved bacterial helicase, has been shown to unwind D-loops and displace RecA single-stranded DNA (ssDNA) filaments, thereby limiting the persistence of recombinogenic intermediates (Lloyd and Sharples, 1993; McGlynn and Lloyd, 1999; Vincent et al., 1996).

When replication forks converge at the *yjhR* locus, the absence of RecG likely allows for enhanced strand invasion and branch migration, promoting recombination events that can lead to the formation of revertants. The significant increase in recombination frequencies and revertants at *yjhR* in *recG*$^{-}$ mutants underscores the importance of RecG in maintaining genome stability by fine-tuning the balance between the formation and resolution of recombination intermediates during fork fusion. The interplay between RecG and other recombination factors at sites of fork convergence, such as *yjhR* in the *oriZ*$^{+}$ background, warrants further investigation.

To further investigate the role of RecG in processing recombination intermediates during replication fork fusion events, we examined the effect of deleting *recG* in a double origin background where the recombination cassette was inserted into the native termination area, at the *narU* locus. This approach allowed us to assess the impact of RecG on revertant formation and fork processing specifically within the native termination area in the context of the replication fork trap. Deleting *recG* in a double origin background showed significant increase in frequency of reverted cells, when measured at *narU-MX4* (Figure 31), implicating RecG in processing intermediates inside the termination area, paired with the MFA results showing there is increased copy number where forks fuse in cells lacking RecG helicase (Rudolph et al., 2010a, 2013). These backgrounds have a fork arrested prematurely at *terC* and therefore also implicate RecG in processing stalled forks at Tus-*ter* pause sites.

The role of RecG in processing intermediates generated at stalled replication forks has been well-established (Briggs et al., 2004; McGlynn and Lloyd, 2000). RecG is known to catalyze the reversal of stalled replication forks, converting them into Holliday junction-like structures that can be processed by various pathways, including recombination (McGlynn and Lloyd, 2000). This fork reversal activity is thought to be important for the stabilization and repair of stalled forks, preventing them from collapsing and generating double-strand breaks (Briggs et al., 2004). In the absence of RecG, stalled forks at Tus-*ter* pause sites within the termination area are likely to persist and potentially collapse, leading to an increase in recombination intermediates and, consequently, a higher frequency of revertant formation. The premature arrest of forks at *terC* in the double origin background further exacerbates this situation,

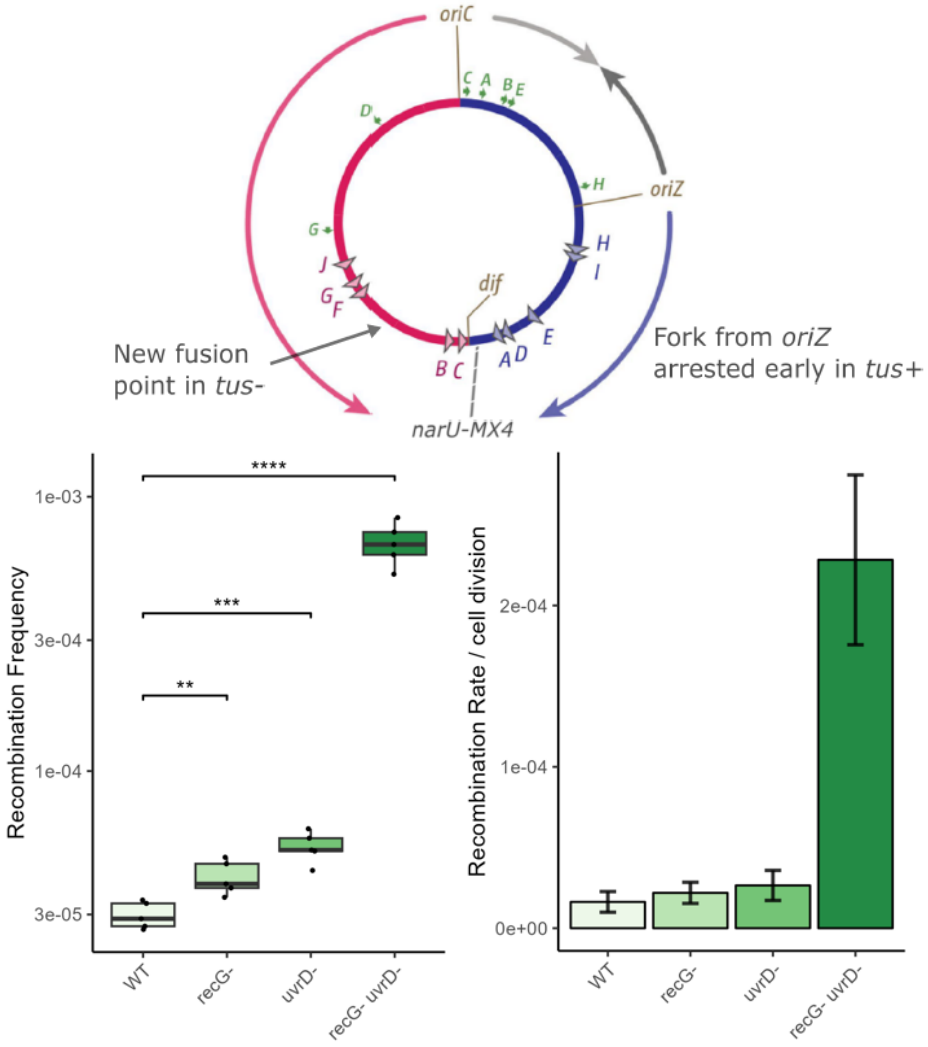

Figure 31: Frequency of *recG* and *uvrD* mutants showing synergistic effect of the two helicases within the termination area. Figure showing 5 individual data points for each strain as frequencies, representative of 5 independent experiments, where rates were calculated at the end of each experiment. Rate bargraphs show average of the 4 rates and error bars showing SEM and ANOVA test was performed on each group of data points to determine p values.

highlighting the critical role of RecG in managing stalled forks and maintaining genome stability in the termination area. These findings suggest that RecG is not only important for processing recombination intermediates during fork fusion events but also plays a crucial role in the rescue and stabilization of stalled forks at Tus-*ter* sites, preventing their collapse and the subsequent generation of recombinogenic structures.

Following the *recG* data, the next question was what effect would UvrD have on recombination inside the termination area. The model includes RecG's role in processing 3' flap structures during fork fusions and with RecG's role in resolving D-loop structures, but does not have a role for UvrD, which has roles reversing stalled RNAP at replication-transcription conflicts, MMR and NER, also being shown to interact with Tus when arresting stalled forks to allow replisomes to overcome the Tus-*ter* barrier (Florés et al., 2005).

The prediction would be that if UvrD is playing a role locally inside the termination area to mitigate issues occurring during fork fusion that may generate recombination structures, then cells lacking UvrD should show increased frequencies, which is what I observed. The high rates in *recG uvrD* double mutants directly implies synergistic effect between the two helicase that can generate over-replication and accumulation of DNA intermediates which are pro-recombinogenic.

The frequency data, visualized in Figure 31, shows that the probability of generating a revertant per cell cycle is increased in cells lacking RecG, when the recombination cassette is placed at *narU-MX4* within the termination area. This effect is expected given the documented over-replication present in the termination area in *recG*$^{-}$ mutants Wendel et al. (2014). Considering the significant over-replication of the termination area in *recG*$^{-}$ cells, the reversion rate data shows only a moderate increase in revertants compared to wildtype when the cassette is at *narU*. This would suggest that simply a higher copy number does not directly translate to higher rates of recombination, although it may contribute (Rudolph et al., 2013). Other factors, such as the availability of recombination substrates and the efficiency of recombination processing, likely play a role in determining the overall recombination frequency.

Interestingly, there is a clear synergistic effect of combining the *recG*$^{-}$ deletion with *uvrD*$^{-}$, as the single mutants alone produce only modest, albeit significant, increases in local recombination. This suggests that RecG and UvrD may have overlapping or complementary roles in suppressing recombination events.

The model for over-replication occurring in *recG* mutants proposes that fork fusion events pro-

duce 3' flaps, which are then processed by PriA to load replisomes and re-replicate the DNA (Dimude et al., 2015; Midgley-Smith et al., 2018; Rudolph et al., 2009, 2013). UvrD has been implicated in the processing of recombination intermediates and the dissolution of RecA filaments (Petrova et al., 2015; Veaute et al., 2005), and the data shown in Figure 31 illustrate that in the absence of both RecG and UvrD, the accumulation of unresolved recombination intermediates and persistent RecA filaments may further exacerbate the over-replication phenotype, leading to the observed synergistic increase in recombination.

Previous studies have shown that UvrD plays a role in regulating recombination at replication forks and preventing the formation of aberrant recombination intermediates (Florés et al., 2005). The loss of UvrD in the *recG* background may further destabilize stalled replication forks and promote inappropriate recombination events. The combined absence of these two key regulators of recombination may explain the significantly elevated recombination frequency observed in the *recG uvrD* double mutant.

Overall, the data is consistent with the model that RecG and UvrD work together to suppress over-replication and aberrant recombination events, particularly in the context of replication fork fusion and stalling, evident by the elevation in the double origin mutant. This work presents the first demonstration of the synergistic effect of RecG and UvrD in regulating recombination at colliding replication forks and Tus-*ter* barriers within the termination region. The synergistic effect of the *recG uvrD* double mutant highlights the importance of these two enzymes in maintaining genome stability by processing recombination intermediates, regulating RecA filament formation, and facilitating replication fork progression through the termination region. This novel finding shown in Figure 31 builds upon the previous work by Bidnenko and colleagues, who showed that UvrD can unwind Tus-*ter* barriers and facilitate stalled fork progression (Bidnenko et al., 2006).

Mechanistic insight and *recA* mutants

From the data obtained from the reversion rate experiments, one question still remained of if we could derive mechanistic insight on how the cassette is reverted. By deleting *recA*, then measuring the rates again compared to *recA*$^{+}$ strains with either a single or double origin, we should be able to see that if all reversion events are dependent on RecA. If this is the case, then single origin and double origin strains should drop to a base level rate that is comparable. We would expect some revertants to be generated in the *recA*$^{-}$ mutant, but we would hypothesize that the frequency to be drastically reduced in this background. The results indeed support the idea of a major role of RecA in reversion, however there appears to also be a RecA-independent mechanism of reversion.

By testing a *recA*$^{-}$ mutant in strains which are *oriC*$^{+}$ *oriZ*$^{+}$ we can see the effect of revertant generation in cells lacking RecA. The data show that the probability of reversion of the *kankanMX4* cassette is significantly lowered in *recA*$^{-}$ mutants, although they are still able to form (Figure 32). The explanation to why reversions can still take place without *recA* are down to template slippage and/or RecA independent mechanisms, possibly involving RarA, which is still present. By testing a *recA*$^{-}$ mutant in strains which are *oriC*$^{+}$ *oriZ*$^{+}$, we can investigate the effect of reversion generation in cells lacking canonical recombination pathways.

oriZ^{+} *recA*$^{-}$ revertants in the *yjhR* background were measured, the rationale being that deleting *recA* should reduce revertant probability if the mechanism is via homologous recombination, which is RecA dependent.

These data indicate that the contribution of fork fusions to recombination rates is reduced in *recA*$^{-}$ mutants, however there are increased rates in the double origin compared to single origin when RecA is absent. From Figure 32 we can see a similar decrease pattern in both *recA*$^{-}$ strains. This comparable decrease implies a RecA-independent pathway for reversion generation, meaning, that increase at fork fusion sites is partly driven by RecA, but there is a RecA-independent pathway in which revertants are still able to form.

These data suggest low level RecA-independent reversion generation is possible. One observation supporting our hypothesis is that *oriZ*$^{+}$ strains have an increased overall reversion frequency compared to wild-type. Another is that *recA*$^{-}$ frequencies are significantly reduced versus wild-type. The fact that there is an increase in the *oriZ*$^{+}$ background when *recA* is deleted essentially indicates a RecA-independent mechanism contributing to reversion at the *kankanMX4*

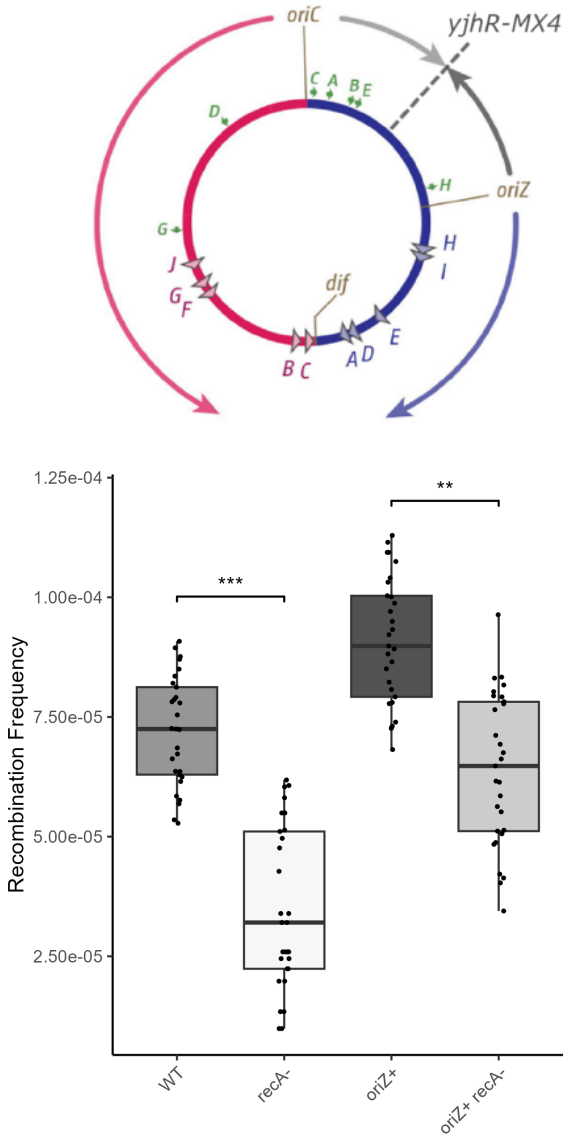

Figure 32: Frequency of *recA* mutants measured at *yjhR* where forks fuse in oriZ^{+} background. The overall reduction when RecA is deleted shows RecA dependent recombination is at play where forks fuse, but this is not the whole story. A RecA-independent mechanism is still at play when forks fuse that lead to reversion of the *kankanMX4* cassette. Figure showing 33 individual data points for each strain as frequencies, representative of 3 independent experiments, where rates were calculated at the end of each experiment. Rate bar graphs show average of the 4 rates and error bars showing SEM and ANOVA test was performed on each group of data points to determine p values. Asterisks indicate significance.

cassette, where forks are fusing. The clear *recA*$^{-}$ effect in both strain backgrounds shows reduced recombination rates. This unexpected result highlights that aspects of replication termination and fork processing remain unknown, requiring further work to fully elucidate the mechanisms involved.

With RarA being noted as being a key player in RecA-independent recombination (Jain et al., 2021a, 2021b) there is the potential for other enzymes to function and allow reversion, and when forks fuse, there is a greater chance of reversion, even without RecA. Seeing as the over-replication MFA data from the Rudolph lab illustrated that over-replication in *recG*$^{-}$ is entirely dependent on the activity of RecA where forks fuse (Rudolph et al., 2013). The next test would have been to test the effect of *rarA*$^{-}$ single and *rarA*$^{-}$ *recA*$^{-}$ double mutants in single and double origin strains, however, due to time constraints this was not possible.

Chapter 2 - Discussion

The termination area has previously been shown to be a so called 'hot spot' for recombination, with early results showing that the forks blocked at Tus-*ter* complexes may lead to homologous recombination by acting as an entry point for RecBCD at double strand ends of the stalled fork (Horiuchi et al., 1994). More recent work has also shown that the incorporation of *ter* sites into a plasmid model is able to mitigate recombination, showing a possible evolutionary pressure for the Tus-*ter* system controlling fork fusions (Hamilton et al., 2023). This previous research has looked into DNA-protein complexes being the sites of recombination, namely Tus-*ter*, however the consequences of free fusing forks themselves have had little focus, and in reality, forks fusing directly at Tus-*ter* barriers appear to be a rare event and forks mostly fuse stochastically within the boundary of the termination area (Dimude et al., 2016; Jameson et al., 2021; Kono et al., 2012; Rudolph et al., 2013). I wanted to test the hypothesis that free fork fusions themselves may lead to the accumulation of substrates which require recombination for resolution.

My results suggest that replication fork fusion events are a major driver of increased recombination rates in the *E. coli* chromosome. I observed significant elevations in recombination at the ectopic *yjhR* and *tldD* loci when additional fork fusion zones were created by integrating ectopic replication origins *oriZ* or *oriX*. This provides *in vivo* evidence that wherever fork convergence occurs, recombination rates increase locally as a consequence, and this is independent of Tus-*ter* complexes. This has interesting implications for eukaryotic cells, which undergo thousands of replication fork fusion events per cell cycle. Eukaryotes likely require highly efficient helicases like Pif1 and Rrm3 to properly regress or dismantle aberrant fork structures and prevent excessive instability from the sheer number of fork collisions (Malone et al., 2022; Steinacher et al., 2012).

Importantly, I found no significant increase in recombination at *yjhR* in *oriC*$^{+}$ *oriX*$^{+}$ strains, where the fork fusion point is unaltered compared to wildtype. This demonstrates the local stimulation of recombination is due specifically to fork convergence itself, rather than indirect effects like enhanced replication-transcription conflicts from the extra origin. My results at the native *narU* locus in the terminus suggest the replication fork trap may function in part to constrain the spread of potentially pathological recombination arising from fork fusion events. Deletion of tus led to only a modest increase in recombination at *narU*, implying the fork trap helps restrict hyper-recombination to a defined region where it can be properly managed.

I observed some differences in recombination between the ectopic loci in *oriZ* versus *oriX* strains. While rates increased at both *yjhR* and *tldD* when situated at an engineered fork convergence point, the magnitude varied between approximately 2.5-3 fold over wildtype. Analysis of *recA*$^{-}$ mutants provided mechanistic insight, revealing both RecA-dependent and -independent pathways contributing to recombination stimulation at fork fusion zones. The residual recombination in *recA*$^{-}$ mutants likely occurs through template slippage or alternative strand exchange mechanisms (Ede et al., 2011; Jain et al., 2021b).

Overall, my work demonstrates that replication fork convergence events significantly increase recombination rates in *E. coli* at specific fork fusion sites through both RecA-dependent and -independent mechanisms. Regulating these potentially destabilizing recombination events is a key challenge cells face to preserve genomic integrity during replication termination.

Copy number and recombination

The initial idea behind these constructs was to measure the rate of recombination in the termination area compared with the rest of the chromosome. One might expect, given that termination area has been previously reported to be a hotspot for recombination, that DG010 (carrying the *yjhR-kankanMX4* cassette) should have a lower rate of recombination than that of DG009 (carrying the *narU-kankanMX4* cassette). The results show that the *yjhR-kankanMX4* construct has a higher rate of recombination than the *narU-kankanMX4* construct ($p = <0.05$). One might speculate that the simplest explanation for the decreased rates at *narU* compared with *yjhR* is that the replication fork trap is mitigating the consequences of replisome fusions, thereby leading to a decreased recombination frequency. However, after familiarization of the system, it appears that copy number at these two distant sites is having a more direct impact on recombination rates, rather than any architectural or site specific function. This is illustrated by measurements taken at *tldD* where the values are more comparable to *yjhR*, being that they are both in the 'northern chromosomal hemisphere'. Therefore, it is prudent to compare recombination sites to one another only if they are likely to have a comparable copy number to each other, which *narU* and *yjhR* do not in WT cells. By inserting the cassette near *tldD*, we have a more direct comparison with *yjhR*, as both are within the same 'hemisphere' of the chromosome, if we were to separate the *E. coli* chromosome into northern and southern hemispheres. By introducing the cassette near *tldD* we can test a third chromosomal location to get a better picture of recombination across the chromosome. Across all three locations, my results show the same pattern, that whenever forks are actively fusing near the recombination cassette then the rates

increase. In *tus*$^{-}$ strains, there is also a mild increase in the frequency of revertants, highlighting that recombination directly at Tus-*ter* barriers is not responsible for the increase.

Fork fusions and reversion rates

The aim of these experiments were to test if reversion frequencies and rates are increased where there is an ectopic fusion site on the chromosome, and not to determine the effect of recombination where forks do not fuse.

Comparing the frequency data for *yjhR* and *tldD*, the pattern is consistently showing that there are higher recombination frequencies in strains which have an ectopic termination event from the introduction of an ectopic origin into that replichore. These data support a model where fork fusions lead to increased reversion of the *kankanMX4* cassette.

The data also shows that there is less recombination taking place when the ectopic termination area is in the opposite replichore. Therefore, *oriZ*$^{+}$ *yjhR* is comparable with *oriX*$^{+}$ *tldD*, which both have increased recombination taking place, and *oriX*$^{+}$ *yjhR* is comparable with *oriZ*$^{+}$ *tldD*, which both have greatly reduced levels of recombination.

It has been shown that *oriX*$^{+}$ strains suffer a reduced growth phenotype, where *oriZ* has only minor delays. Therefore, the assumption was originally that the drop in revertant colonies in *oriX*$^{+}$ *yjhR* cells could be caused by the reduced fitness in this strain. Testing the recombination levels in *oriZ*$^{+}$ *tldD* would show us if it is the *oriX*$^{+}$ strains themselves that are causing the reduced levels of recombination, or if there was a previously unknown effect occurring from disruption in the chromosomal symmetry. The comparable drop in recombination of *oriZ*$^{+}$ *tldD* shows that it is indeed that latter and levels are not comparable to WT.

Initially, we might predict that if fork fusions are directly contributing to the frequency of recombination, then the observed increases at *yjhR* and *tldD* should be comparable. However, when looking at the raw rate of recombinations, we see there is a local difference between the rates. The simplest explanation is that rates at are determined locally and represent that particular chromosomal region under the conditions measured. For instance, one might predict a slightly higher rate of recombination in the right hand replichore compared with the left due to the presence of the *rrnCABEH* cluster. These highly transcribed operons could cause a barrier for replisomes approaching from the wrong direction, leading to head on collisions between RNAPs and replisomes, which may result in fork arrest and require recombination dependent replication restart However, despite local differences in chromosomal location, both locations point to a fusion-dependent increase in reversion rates at the *kankanMX4* cassette and the same trend was observed in different chromosomal locations.

oriX Strains and General Fitness

When comparing the frequencies there was a 2.5 fold and or 2.8 fold increases at *tldD* and *yjhR*, respectively, illustrating that the level of recombination varies across different points along the chromosome and disrupting the natural chromosomal architecture has varying degrees of effects across the chromosome. For instance, not all *oriZ*$^{+}$ strains have a drastic increase in recombination compared to the WT, showing that disrupting the chromosome in general with additional origins do not lead to global increased recombination rates.

When additional origins are included into the chromosome things can escalate quickly inside the termination area, with the measurements at *narU* in an *oriZ*$^{+}$ background showing 2 fold increase in recombination rate compared to a single origin. The *oriX*$^{+}$ constructs required 2 days for pin prick colonies to develop, showing the reduced fitness of the *oriX*$^{+}$ strain in general. *oriZ*$^{+}$ *tus*$^{-}$ also required two day incubation periods, showing that without the replication fork trap, the level of recombination may lead to a reduced fitness when compared with *oriZ*$^{+}$ strains in which Tus is present. There is likely a dual effect when measuring the effect of Tus. When Tus is present and a extra origin present, there will indeed be recombination of stalled forks held up at Tus-*ter* complexes. However, in a *oriZ*$^{+}$ *tus*$^{-}$ strain, we see a slight increase in recombination at *narU*, showing recombination is not dependent on forks held at Tus. The resulting recombination rate is the therefore the product of recombination caused by the absence of Tus minus the recombination normally caused from forks at Tus.

At first glance one might assume that reduced number of mutants in *oriZ*$^{+}$ and *tus*$^{-}$ single mutants that more recombination is occurring in a combining strain of the *oriZ*$^{+}$ *tus*$^{-}$, however the slow growth phenotype on antibiotic goes against this idea. The situation is therefore nuanced and requires proper interpretation which marries the rate of recombination and fitness of all strains using this system to come to appropriate conclusions. Even after 2 days growing, the amount of colonies for DG027-032 were less than DG009 (WT *narU-MX4*). DG033 (*oriZ*$^{+}$ *tus*$^{-}$) had a significantly higher frequency of mutants compared with WT, but less than *tus*$^{-}$.

It is possible that the location of the *oriX*$^{+}$ insertion is the cause of the observed slower growth phenotype. As kanamycin works better with Mu medium instead of LB, this higher salt dependency for this system may have made *oriX*$^{+}$ strain more sick compared to WT or even *oriZ*$^{+}$. For this reason, I did not test any combinations of *recG*$^{-}$ or *recA*$^{-}$ in the *oriX*$^{+}$ background as strains always take twice as long to see colonies, meaning the longer growth time could mean reduced fitness.

Although the increase seen in *oriZ*$^{+}$ strains mimics that seen in single origin strains, there was also longer incubation times needed for certain *oriZ*$^{+}$ cells. The somewhat lowered fitness of *oriZ*$^{+}$ *recG*$^{-}$ cells may predict that there would be less viable cells with the reverted cassette, however even with this reduced fitness and increase doubling time, the frequency of revertants are higher in *recG*$^{-}$ compared with *recG*$^{+}$ cells with an additional ectopically placed origin. Although these data were not meant to assess fitness, the obvious small colony size of some strains may indicate lack of fitness through their increased incubation time. This was not as drastic with *oriZ*, compared to *oriX*, but still worth noting.

The effect of deleting *recG* and *uvrD*

The previously published marker frequency analysis from the *recG*$^{-}$ mutant results show a notable increased copy number in the termination area (Rudolph et al., 2013). The recombination rate data resented here corroborate these findings to some degree, although the MFA data clearly show a strong peak of over-replication. This is a significant finding as it suggests that the RecG protein plays a crucial role in controlling replication and recombination within this region. The termination area is a critical part of the DNA replication process, where the replication of the DNA molecule ends (Rudolph et al., 2013; Rudolph et al., 2019). In a normal scenario, the RecG protein would be expected to regulate this process, ensuring that replication does not exceed the necessary levels. However, in the absence of RecG, as seen in *recG* mutants, there appears to be an uncontrolled increase in replication. This could potentially lead to an overproduction of certain genes, which may have various implications for the organism's overall genetic function and stability.

Interestingly, despite the increased replication in the termination area observed in *recG* mutants, there was only a modest increase in reversion frequencies and rates as part of this project. This finding is somewhat counterintuitive as one might expect that an increase in replication would correspondingly lead to a higher rate of reversion. Reversion, in this context, refers to the process by which a DNA sequence returns to its original state after a mutation has occurred. The modest increase observed suggests that while the absence of RecG does lead to more replication, it does not significantly affect the rate at which these replications revert to their original state.

Based on these observations, one might predict a drastic increase in reversion frequencies and rates in the termination area without RecG. However, the data does not support this prediction as only a modest increase in reversion frequencies and rates were observed. Despite the increased aberrant initiation taking place inside the termination area, my data suggests that simply increasing copy number does not always result in comparative recombination rates at the same locus. This highlights the complexity of genetic processes and underscores that multiple factors beyond just copy number and deleting one helicase, such as RecG, can influence recombination rates. It also suggests that other proteins or mechanisms may be at play in regulating reversion frequencies and rates, even in the absence of RecG. This finding opens up new avenues for further research into the intricate processes governing DNA replication and reversion, for instance looking into the level of expression of other helicases and nucleases that can

process intermediates known to stimulate over-replication as well as recombination (Dimude et al., 2015; Midgley-Smith et al., 2018; Midgley-Smith et al., 2019).

Previous work from our lab has shown that deletion of the Tus terminator protein or the RecG helicase results in increased copy number and over-replication of the termination area of the *E. coli* chromosome (Dimude et al., 2015; Rudolph et al., 2013). This over-replication is thought to occur due to the accumulation of aberrant DNA structures when replication forks converge, such as 3' flaps, that are normally processed by RecG and 3' exonucleases (Dimude et al., 2015; Midgley-Smith et al., 2018; Midgley-Smith et al., 2019).

One likely explanation for the discrepancy is that the absence of RecG impairs DNA repair and homologous recombination pathways genome-wide, counteracting the effect of having more sister chromatids available (McGlynn and Lloyd, 1999; Midgley-Smith et al., 2018; Rudolph et al., 2013). Thus other helicases and recombination proteins cannot fully compensate for the loss of RecG even though substrate for recombination has increased. This highlights the importance of RecG's functions in processing branched DNA structures like replication forks and D-loops to facilitate homologous recombination. Further work will be needed to fully delineate the relationships between copy number, aberrant DNA structures, and recombination rates at defined chromosomal regions.

Reversion rates in an *oriX* and *oriZ* background

I measured rates of recombination at the *narU*-kanMX4 locus in strains with and without an active replication fork trap system. As a control, we also assayed recombination at a kanMX4 cassette integrated at the *yjhR* locus. We initially observed that inactivation of the RFT resulted in increased recombination at *narU* compared to WT, suggesting the RFT suppresses recombination in WT cells.

These results demonstrate that fork fusion in an ectopic region of the chromosome promotes the reversion of the *kankanMX4* cassette at the convergence point restoring kanamycin resistance. Disruption of replichore symmetry increases recombination rates by both fork fusion-dependent and -independent mechanisms, such as those mediated at Tus-*ter* blocks. The RFT may facilitate replication restart by homologous recombination when forks block at Tus-*ter* sites requiring restart, while also coordinating proper chromosome segregation.

Recombination rates and the replication fork trap

Here, we measured rates of recombination at the *narU-kanMX4* locus in strains with and without an active replication fork trap (RFT) system mediated by Tus and *ter* sites. This builds on previous work from our lab showing increased copy number and over-replication in the termination area in recG and tus mutant strains. We hypothesized that recombination rates might also increase in these backgrounds due to the presence of excess 3' flap structures that could engage in homologous recombination to restart stalled or collapsed replication forks at Tus-*ter* barriers.

To test this, we constructed strains with the native RFT system or with the *tus* gene deleted to inactivate the RFT. We analyzed strains with single origins and thoses containing an additional ectopic origin, *oriZ* or *oriX*, which disrupts the symmetry of the two replichores.

The data from the double origin experiment provides compelling evidence that fork fusions can indeed instigate recombination. This leads us to question the role of the Replication Fork Trap (RFT) in mitigating these effects. According to Horiuchi's 1994 study, the absence of Tus-*ter* formation should theoretically result in a decrease (Horiuchi et al., 1994).

However, the observed increase in *tus*$^{-}$ aligns with the notion that the RFT is indeed playing a role in regulating recombination. While this supports the idea of RFT's involvement, it's important to consider that there may be other factors or mechanisms at play contributing to this observed increase, likely from the fork trap working to reduce recombination and maintain local stability of the area when forks fuse. This data is in line with (Hamilton et al., 2023) results where they showed a reduction in recombination when *ter* sites were added to their plasmid model. Without Tus present, recombination at Tus simply would not happen so the increase in *tus*$^{-}$ cells cannot be due to the activity of recombination taking place at stalled forks held at Tus-*ter*.

The rates were higher in double origin strains, where *tus*$^{+}$ cells had the highest rates of recombination, in line with early work by Horiuchi in which recombination is triggered from stalled forks held at Tus-*ter* barriers. The lower rates in a *oriZ*$^{+}$ *tus*$^{-}$ background suggest forks now fuse well into the right hand replichore away from the normal fusion point, which reduces recombination and probability of generating a revertant, near *narU*.

Here, we suggest that the termination area is more susceptible to changes in recombination rates when the chromosomal architecture is disrupted with the addition of ectopic origins.

tus^{-} mutants show a slight increase in recombination compared with the WT, whereas in *oriZ*$^{+}$ strains there are increased rates. , and also early models which show Tus-*ter* barriers are sites of recombination through potential entry sites for RecBCD (Horiuchi et al., 1994; Sinha et al., 2020). The results that rates are slightly increased in *tus*$^{-}$ cells go against the idea that Tus-*ter* barriers are causing significant recombination on their own. My data shows there is another effect here of the fork trap seeming to lower rates of recombination in single origin strains.

These results demonstrate a complex relationship between replication fork blocking, replichore symmetry, and recombination. The RFT likely evolved to serve functions beyond simply suppressing recombination, and naturally further work is needed to elucidate the precise mechanisms connecting RFT function, replication fork dynamics, and recombination in the termination area as evolutionary drivers.

By testing the *oriZ*$^{+}$ strains measured at *narU*, we set up the situation as described by Horiuchi (Horiuchi and Fujimura, 1995) whereby recombination is able to gain access to the double strand end at a stalled fork if, for example, the lagging strand is nicked, at a stably arrested fork at Tus-*ter* barrier, in this case at *terC*. When the fork trap is inactive in this background, forks are permitted to move through the termination area, and as measured by marker frequency analysis, forks now fuse well into the left hand replichore and away from *narU*, resulting in lower rates compared with *tus*$^{+}$ backgrounds, in line with the hypothesis that fork fusions create intermediates which require recombination for repair/maintenance of chromosome stability (Dimude et al., 2015; Hamilton et al., 2023; Midgley-Smith et al., 2018; Rudolph et al., 2010c).

Interestingly, there are two ways to interpret the *tus*$^{-}$ data, for and against the idea of recombination directly at Tus-*ter* barriers. The data show a slight increase in revertant frequency at *narU* in the *tus*$^{-}$ background. The increase would go against the notion that forks held up at Tus are sites of recombination, simply because there isn't an increase. Whether or not any alterations in Tus-dependent recombination would be measured by this assay, is hard to say, especially given that forks fusing at Tus is not a common event (Jameson et al., 2021). The increase in rates at *narU* in *oriZ*$^{+}$ likewise, does not give an explanation for why the recombination rates are elevated in the termination area with an ectopic additional origin. The increased rates seen in this background, where a fork will be stably arrested prematurely at *terC*, are likely a combination of more passive RecBCD activity at stably arrested forks, and more active DSB generation from nascent forks approaching from *oriZ*, as well as other stochastic features inside the area.

RecA-independent mechanisms of homologous recombination

The observation that reversion rates increase at *yjhR* in *oriZ*$^{+}$ strains even without RecA suggests a RecA-independent mechanism contributes to instability when replication forks converge. To elucidate this mechanism, strains with deletions of other recombination genes like *radA*, *rarA*, _recBCD_and *ruvABC* could be constructed and analyzed to determine if they are required for the residual reversion events. Examining mutations in *priA*, topoisomerases, and gyrase could provide insight into whether replication restart, supercoiling changes, or aberrant fork structures like flaps might promote template slippage when forks meet. Inserting *ter* sites surrounding *yjhR* in the *oriZ*$^{+}$ *recA*$^{-}$ background would reveal if blocking fork convergence suppresses instability without RecA present. Finally, DNA combing could allow direct visualization of fork progression and fusion dynamics in the absence of RecA. Determining the processes enabling chromosome duplication and segregation when RecA-mediated repair is impaired will provide fundamental insight into replication termination mechanisms and how cells resolve problems arising at fork convergence points. This work has uncovered a gap in our knowledge of how chromosome maintenance is coordinated at termination, warranting further study to elucidate the interplay between replication completion, recombination, and other DNA metabolic pathways needed to preserve genome integrity during DNA replication (Jain et al., 2021b).

RarA was shown to function in intermolecular recombination between plasmids, particularly involving short homologies under 200 bp. My recombination reporter cassette contains 266 bp direct repeats, within the size range where RarA makes a substantial contribution. RarA may act at stalled forks to promote strand invasion and D-loop formation independently of RecA and can form 3' flaps from duplex DNA (Jain et al., 2021b). The ATPase activity of RarA could provide the energy for these events. If RarA can invade homologous DNA and generate branched intermediates, this would fit into our model of 3' flap generation from converging replication forks Dimude et al. (2016).

The findings from the Cox lab are relevant when considering the increased reversion rates I observe at the ectopic *yjhR* locus in *oriZ*$^{+}$ *recA*$^{-}$ mutants (Jain et al., 2021a, 2021b; Stanage et al., 2017). However, their results point to proteins like RarA that could promote RecA-independent recombination intermediates even without RecA present. RarA may act at stalled forks to invade homologous DNA and generate branched joint molecules. I plan to test this model by constructing *oriZ*$^{+}$ *rarA*$^{-}$ and *oriZ*$^{+}$ *rarA*$^{-}$ mutants, followed by *oriZ*$^{+}$ *rarA*$^{-}$ *recA*$^{-}$ triple mutants and

measuring recombination rates at *yjhR-kankanMX4*. This would determine if RarA facilitates the residual events detected.

Jain and colleagues (Jain et al., 2021b) investigated RecA-independent recombination systematically in *E. coli* using a two-plasmid assay. By measuring recombination between plasmids containing variable homology lengths, they showed substantial RecA-independent crossing over occurring even at very short homologies under 50bp.

Jain and colleagues also discussed the potential role of RarA, a protein involved in DNA repair and recombination, in RecA-independent recombination. They suggested that RarA might assist in the generation of a 3' flap during the convergence of replication forks. This 3' flap could serve as a substrate for the recombination machinery, facilitating the exchange of genetic material between the two DNA strands. This hypothesis aligns with previous studies that have shown RarA's involvement in processing and stabilizing stalled replication forks, which could potentially lead to the formation of these 3' flaps. However, further research is needed to confirm this proposed function of RarA in RecA-independent recombination as forks converge to complete DNA replication.

The results from our study indicate that even in the absence of RecA, an essential protein for homologous recombination, there is still an increase in recombination at the sites where replication forks converge. This suggests the existence of a RecA-independent pathway that contributes to recombination during replication termination. One plausible explanation for this observation is the role of RarA, a protein known to be involved in the repair of collapsed replication forks. RarA is recruited to the site of the broken fork through its interaction with SSB proteins, where it binds to the double-stranded DNA end on the lagging strand arm. Utilizing the energy of ATP binding and hydrolysis, RarA locally separates the DNA strands at the end, creating a single-stranded flap (Stanage et al., 2017). This single-stranded flap provides a suitable substrate for the replicative helicase DnaB to load onto and continue unwinding the parental duplex DNA, thereby allowing replication to proceed downstream of the discontinuity site without the need for replisome disassembly and reassembly.

In the termination area, where stalled forks may accumulate ssDNA intermediates, tight regulation of exonucleases is likely critical to prevent pathological RecA-independent events. The replication fork trap architecture may have evolved specifically to spatially restrict unregulated recombination caused by improper fork collisions where ssDNA could accumulate. By containing these potentially highly mutagenic events to a defined zone, the fork trap helps maintain

overall genome stability. In this context, the strand separation activity of RarA generates a viable DNA substrate for the replication machinery to continue synthesizing DNA, effectively allowing the fork to restart after collapse without RecA-mediated homologous recombination playing a major role (Stanage et al., 2017). However, without RecA to form filaments on this ssDNA flap, the canonical recombination pathway cannot occur, but 3' flaps may still be able to be formed from activity of RarA (Jain et al., 2021b).

Our findings thus point towards a role of RarA in RecA-independent recombination specifically linked to replication fork fusion events, although we do not have any supporting data for this. Further investigation into this pathway could provide valuable insights into the proteins enabling template exchanges and slippage in the absence of RecA, revealing backup pathways that preserve chromosome integrity when homology-directed repair is limited. This work has uncovered an intriguing aspect of recombination-independent fork processing that warrants further molecular dissection. Understanding how cells resolve termination intermediates and complete replication without full recombination capacity will shed light on fundamental backup systems that maintain genome stability during DNA replication.

Reversion Rate Methods and Optimizations

A key factor is that my system has an extremely high basal mutation rate, with mutants frequently in the hundreds or thousands per culture. This violates assumptions of low mutation rates in classic methods like the Lea-Coulson median and makes maximum likelihood estimations numerically unstable. Accumulation of revertants prior to plating also skews results if not accounted for.

To improve accuracy and precision, I increased culture volumes to 11 mL and the number of parallel cultures to 11. This allowed capturing the full range of mutant distributions. I also incorporated measuring the viable cell titer in each culture to account for differences in growth. For analysis, I utilized the R package Flan which implements more advanced statistical methods like the Ma-Sandri-Sarkar maximum likelihood estimator.

Compared to older protocols, this updated approach provides better fitting of the observed mutant distribution to calculate mutation rates. Key advantages are the ability to handle "jackpot" cultures with very high mutant numbers, capturing potential culture-to-culture variability, and accounting for differences in final cell densities between replicates.

For the specific kankanMX4 system, additional modifications were made to reduce accumulation of revertants before fluctuation tests. Despite high basal rates, the optimized methodology enables reliable quantification of how factors like additional origins or helicase mutations modulate recombination. While historical methods work well for lower mutation rates, adapting protocols and analysis tools significantly improved data quality for my hyper-recombinogenic system.

The original fluctuation analysis method introduced by Luria and Delbrück in 1943 made several simplifying assumptions, including no pre-existing mutants in the inoculum and constant final cell counts across replicate cultures. Subsequent studies in the 1990s and 2000s revealed poor fits between the theoretical models and experimental data, identifying issues like jackpot cultures and accumulation of revertants prior to plating. This led to the development of new statistical methods and tools aimed at addressing limitations of the classical framework. Other innovations included webtools like bz-Rates in 2015 that incorporated differences in growth rates between mutants and wild-type. Recent R packages like rSalvador and flan have implemented probability generating functions to deal with unrealistic assumptions. Determining optimal sample size and the use of partial plating to reduce needed replicates have also been exam-

ined. The overarching theme is a move towards statistical methods and tools that better reflect the biological realities of each experimental system. The latest advance is inferring recombination rates entirely from large sequencing datasets, suggesting a possible future direction.

In summary, the decades since Luria and Delbrück's foundational work have seen extensive refinements to fluctuation analysis methodology. While the basic experimental procedure remains similar, new statistical approaches have emerged to address limitations related to assumptions around pre-existing mutants, equivalent growth rates, plating efficiency, and appropriate sample size. Maximum likelihood estimators are preferred for high mutation rate scenarios, while probability generating functions help account for plating biases. Overall, choosing the analytical method that best captures the specifics of each experimental system appears crucial. Ongoing innovations in areas like machine learning applied to sequencing data may further expand the toolkit for obtaining robust mutation and recombination rate estimates via fluctuation analysis.

The classical fluctuation assay and its statistical analyses have gone through many iterations, with new tools and methods addressing limitations like jackpots, variable final cell counts, differences in growth rates, and pre-existing mutants. However, most methods still rely on assumptions like full plating efficiency. The maximum likelihood approach has emerged as a preferred method for high mutation rates, but newer probability generating function methods account for issues like plating efficiency. Reducing needed replicates via partial plating has been suggested. Overall, choosing the method that best fits the realities of the biological system remains key. But fluctuation analysis remains in use today, with ongoing refinements to address its limitations.

Timeline of Key Advancements in Fluctuation Analysis:

In the early days of fluctuation analysis, introduced by Luria and Delbrück in 1943 (Luria and Delbrück, 1943), the median method of calculating mutation rates from culture replicates was considered sufficient. The assumptions made, such as no pre-existing mutants in the inoculum and full plating efficiency, were reasonable approximations that allowed rough estimates of mutation rates to be determined. However, as the field progressed, it became clear there were inefficiencies and limitations to this basic approach.

Boe in 1994 (Boe et al., 1994) showed poor agreement between theory and experiment, identifying issues like jackpot cultures and the presence of pre-existing mutants that violated the

assumptions. This demonstrated the median calculation method was not capturing the full complexity of the fluctuation process. In 2000, Rosche and Foster (Foster, 2000) compared methods and recommended the MSS maximum likelihood, which was better at handling high mutation rates. Their key insight was that choosing a method that best reflected the reality of the experiment was crucial for accuracy.

Foster in 2006 (Foster, 2006) stated that no existing method properly accounted for varying final cell counts between replicates. He recommended the MSS maximum likelihood as the current gold standard approach. However, limitations still existed. Ede et al. in 2011 (Ede et al., 2011) used the assay for measuring yeast recombination rates, acknowledging pre-existing mutants as an ongoing issue. In 2012, Hamon and Ycart tried to improve on limitations by developing new statistical methods (Hamon and Ycart, 2012).

The field responded by moving towards methods like the maximum likelihood employed by Krasovec in 2014 (Krašovec et al., 2014), which allowed better comparisons of mutation rates between replicates. Gillet-Markowska in 2015 (Gillet-Markowska et al., 2015) developed the bz-Rates tool that incorporated differences in growth rates. New software tools like rSalvador and flan were created to implement the latest methods. Zheng in 2017 (Zheng, 2017) examined sample size optimization, suggesting partial plating could reduce the needed number of replicates.

My work focuses on calculating recombination rates, which are orders of magnitude higher than point mutation rates. Therefore, it is logical that the methods require adjustment, such as avoiding full saturation and using defined cell densities before plating. Lin and colleagues (Lin and Kussell, 2019) developed a sequencing-based method to estimate recombination rates, which represents the most accurate approach. For my specific application, implementing a similar sequencing method would likely provide the best recombination rate estimates. Overall, the field has advanced from rough approximations to an increasingly nuanced understanding of the intricacies of fluctuation analysis.

Future Directions

The findings of this study suggest that fork fusion events specifically stimulate recombination at convergence points. In particular, we observed an increase in recombination at the *yjhR* locus in the *oriZ*$^{+}$ strain, which is likely due to the creation of a novel fork fusion zone between *oriC* and *oriZ*. This raises an intriguing question: Could the installation of an ectopic replication fork trap at this location mitigate the increased recombination? To test this, we could integrate Tus-*ter* barrier sites flanking the *yjhR* cassette in an *oriZ*$^{+}$ strain and measure recombination rates. Our hypothesis is that blocking fork convergence would suppress the stimulation of recombination at *yjhR*.

In light of these findings, my future work could delve into exploring the role of RecA-independent recombination in recombination rates of local chromosomal regions. A logical starting point would be to test levels in *rarA*$^{-}$, correlating with *recA*$^{-}$ single mutants. Following this, we could construct a *rarA*$^{-}$ *recA*$^{-}$ double mutant and assay recombination rates at *yjhR-kankanMX4*. The prediction is that residual recombination would decrease further if RarA can no longer generate D-loops on which slippage depends.

In summary, by combining these two avenues of investigation, we can not only understand the specific role of fork fusion events in stimulating recombination but also explore the potential of RecA-independent recombination in local chromosomal regions. This approach will allow us to test novel strategies for mitigating increased recombination and provide insights into the underlying mechanisms of recombination.

Overall, further elucidating the relationship between fork trap function, fork fusion events, and varying recombination proteins involved during fork fusion will be an important goal for future studies on how fork fusion events have influenced the evolution of chromosome architecture to maintain genome stability and reduce negative consequences of fork fusion events in *E. coli*.

Chapter 3: The Role of R-loops in Chromosome Dynamics

Background

This chapter examines the level of R-loops in a variety of *E. coli* mutants through dot blot analysis using the S9.6 primary antibody (Vlachos-Breton and Drolet, 2022). A key piece of evidence against the idea of initiation from R-loops in *recG* cells was that overexpressing RNase HI in *recG* mutants did not suppress their over-replication phenotypes (Dimude et al., 2015). Additionally, expressing yeast RNase HI complemented the growth defect of *rnhA* mutants but had no effect on *recG* cells. This strongly implies that R-loops are not the source of SDR in recG cells, contrasting the model put forth by Hong et al. and Kogoma (Dimude et al., 2015; Hong et al., 1995; Kogoma, 1997; Rudolph et al., 2010a).

Instead, the authors found *recG* mutants require RecBCD recombinase activity and RuvABC for their hyper-replication, indicating repaired fork fusions stimulate SDR (Rudolph et al., 2009, 2010a). Work by Wendel and colleagues provided additional evidence for fork fusion-mediated replication in recG cells. They showed RecG processes reversed forks and its loss leads to unscheduled fork fusion events (Wendel et al., 2014). Replication-transcription collisions were also found to threaten genomic stability in cells lacking RecG or RNase HI. This likely occurs due to conflicts between replication induced from SDR and normal transcription. These collisions require processing by RecBCD and RuvABC (Wendel et al., 2020) and shows the role of recombination in completing replication (Azeroglu et al., 2016; Wendel et al., 2014)

Dimude and colleagues in the Rudolph lab clearly demonstrate that the over-replication phenotypes of *recG* and *rnhA* mutants, while superficially similar, actually result from deficiencies in processing different nucleic acid intermediates. R-loops initiate SDR in *rnhA* cells while repaired fork fusions stimulate synthesis in recG mutants (Dimude et al., 2015). This work disproved

a central model in the field and provided a more nuanced perspective on the mechanisms underlying origin-independent replication. Subsequent studies have further solidified fork fusion-mediated replication as the source of SDR in *recG* cells (Drolet and Brochu, 2019; Midgley-Smith et al., 2018). The presented *in vitro* data raise the important question of what role RecG plays *in vivo*. Further investigation into the physiological function of RecG will be needed to elucidate its contribution to genome maintenance processes in the cell.

Validation of R-loops in *E. coli* by RNaseH treatments

In the pursuit of understanding the intricate workings of RNases and their role in R-loop degradation, we embarked on an experiment using wild type gDNA extracts. We subjected these extracts to an hour-long incubation with either sterile water, RNase A, RNase III, or RNase H. The control group was treated with sterile distilled water, while the gDNA concentration was maintained at 100ng/uL.

RNase A, an endonuclease, is known for its ability to cleave single-stranded RNA into 3' phosphoryl and 5' hydroxyl oligonucleotides. On the other hand, RNase H, another endonuclease, specifically targets and degrades the RNA strand of RNA-DNA hybrids, leaving the DNA strand untouched. This unique capability makes RNase H instrumental in removing R-loops and Okazaki fragment RNA primers during DNA replication.

In contrast, RNase III, a dsRNA-specific endonuclease, cleaves double-stranded RNA and some RNA-RNA hybrids. The end products possess 5' phosphate and 3' hydroxyl termini with 2-nucleotide 3' overhangs. This enzyme plays a crucial role in RNA processing and maturation of rRNAs, mRNAs, and small regulatory RNAs.

Our dot blot analysis revealed that after a 1-hour incubation at 37 degrees C, the signal vanished following RNaseH treatment. This finding is consistent with RNase H's known ability to degrade R-loops, shown in FIgure 33

In light of this, we decided to test other RNases as well. Our aim was to ensure that the signal we observed was not an artefact but a true reflection of R-loop presence. By doing so, we hope to provide a comprehensive understanding of the role of different RNases in R-loop degradation. The results of these investigations will be presented in the following section.

The signal is partially effected by RNase III treatment and only moderately affected by RNase A treatment, showing our signal for the S9.6 antibody is specific to the antibody binding to R-loop structures, although the slight drop in signal from SDW and RNase A shows that there are still some RNA structures, likely RNA:RNA structures, which could be leading to a signal boost, showing the importance of using RNase A during the gDNA extraction process, see methods. The signal was moderately reduced in the RNase III treatment steps, also validated with other mutants, so RNase III treatment was not included in all subsequent gDNA extractions to reduce any confounding effect RNase III would have on the R-loop signal. RNase A was included in all subsequent steps to reduce any confounding effect ssRNA would have in the signal coming from

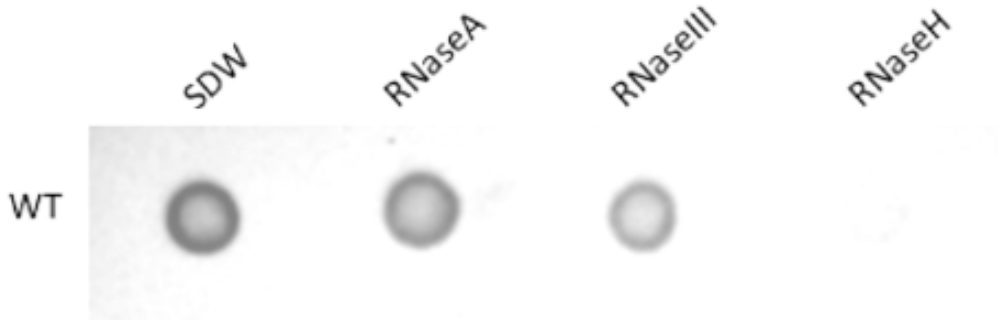

Figure 33: *in vitro* RNaseH treatment showing reduced S9.6 signal when RNaseH is added to gDNA noramalized to 100ng/uL. High concentration was used in order to see strong signal in WT cells. Water and RNase A treatment shows minimal signal change, whereas RNase III shows reduced signal and treatment with RNase H abolishes the signal in WT extracts, as well as other strains (not shown).

the S9.6 antibody, although this was mainly included to reduce noise when measuring gDNA concentrations and making up dilutions to the same concentration.

Following determining these conditions, titration was performed from 100ng, 40ng, 20ng and 10ng to identify the lowest possible concentration to use in order to detect R-loops in WT cells. The WT blots (not shown) indicated too much signal in the 100ng and 40ng, whereas resolution was lost at 10ng, so I determined that all further samples were to be normalized to the concentration of 20ng/uL.

R-loops at different phases of bacterial growth

The gDNA extracted from overnight cultures yielded comparable concentrations of DNA, with a reduced R-loop signal whereas gDNA from exponentially growing cells probed for R-loops and signal was seen in concentrations as little as 5ng/uL, whereas stationary cell extracts required higher concentrations to reveal the R-loop signal, between 10-20ng/uL. Across the board, all mutants showed the same decrease in R-loop signal from overnight cultures. Small signals can still be seen for *rnhA* single mutants in stationary phase gDNA extracts, although to obtain a similar signal we would need to use a higher concentration of DNA, therefore extracts from exponentially growing cells was chosen to get the best possible contrast between mutants.

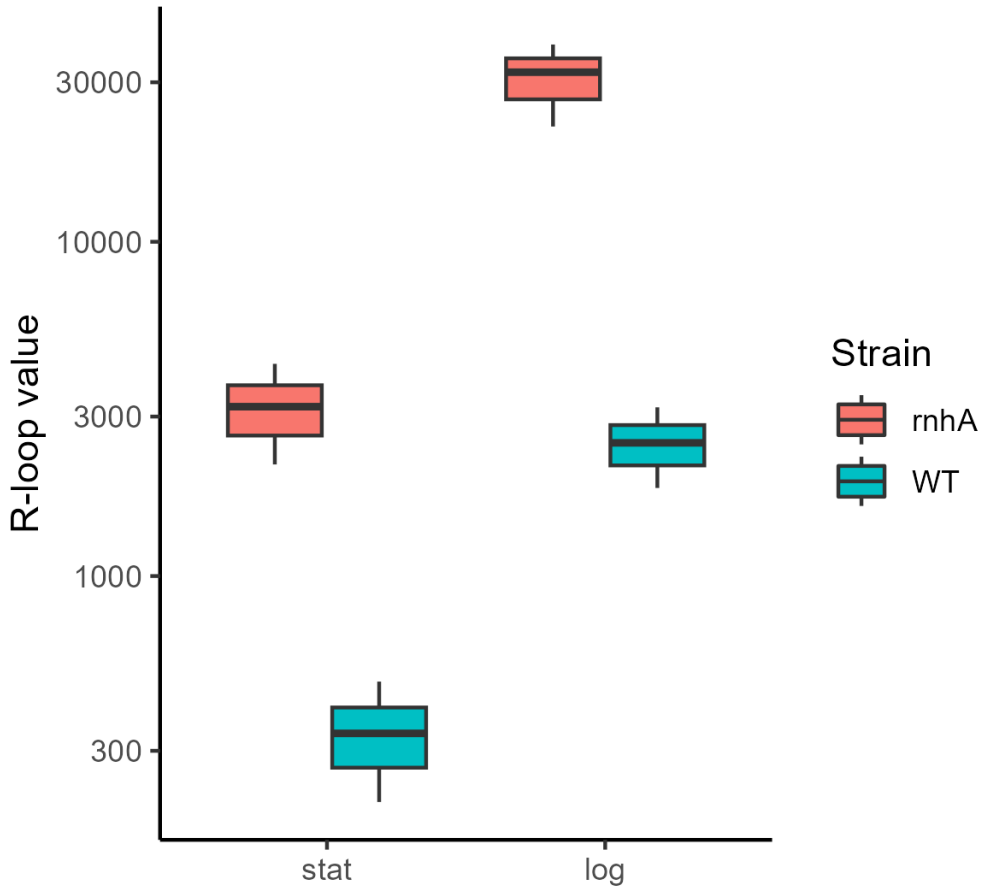

Figure 34: shows the initial data comparing WT and *rnhA* cultures. Naturally, WT is showing the least intense signal, whereas the *rnhA* mutant shows the strongest signal. This is what we would expect and therefore designates WT essentially as the negative control and *rnhA* as the positive control.

I expected to see higher levels of R-loops in exponentially growing cultures compared with sta-

tionary cultures, and indeed levels of R-loops were significantly higher in logarithmically growing cells compared to stationary phase (Figure 34).

Measuring R-loops in log phase cells provides insight into R-loop dynamics during periods of high replication and gene expression, when replication fork fusion events are actively occurring. In contrast, R-loops measured in stationary phase represent an accumulation over time that persists in the absence of key processing enzymes like RNase HI and RecG. While longer-lived R-loops may be more likely to induce aberrant replication initiation, the rapid lethality of the *rnhA recG* double mutant suggests R-loop toxicity occurs quickly during growth. Overall, both growth phases offer merits for studying R-loops, but measurements in log phase may better capture toxic effects arising from R-loop accumulation.

The rapid decrease in R-loops across mutants entering stationary phase provided justification for measuring R-loops at an A_{600} of 0.4 in the experiments here. Further study is still needed to determine if R-loop accumulation is primarily driven by initial loads versus gradual build-up over generations and which growth state best measures R-loops.

The role of RecG in reducing acute R-loops

With the documented role of RecG *in vitro*, one might ask what is the role of RecG *in vivo*? Early work uncovered a role for the multifunctional RecG helicase in modulating R-loop levels in Escherichia coli (Fukuoh et al., 1997; Vincent et al., 1996). R-loops are three-stranded nucleic acid structures containing an RNA-DNA hybrid that readily accumulate in cells lacking RNase HI, which specifically degrades the RNA component of R-loops. Uncontrolled R-loop accumulation leads to pathological over-replication events and genome instability from cSDR (Dimude et al., 2015; Gowrishankar et al., 2013; Rudolph et al., 2010a).

While previous work did not detect increased R-loop levels in *recG* single mutants compared to wild-type cells (Raghunathan et al., 2019), this data of quantitative dot blots showing a modest but statistically significant elevation of R-loops in *ΔrecG* mutants. Notably, treatment with purified RNase HI abolishes this RNA-DNA hybrid signal, indeed implying R-loops are the detectable structure in this assay. From figure 35 the increase in *recG* cells is modest but significant (*p=0.02*) across 5 experimental repeats.

These data indicate RecG does play some role in unwinding RNA-DNA hybrids *in vivo*. The basis for this effect likely stems from RecG's capacity to unwind R-loops using its robust DNA helicase activity (Fukuoh et al., 1997; McGlynn and Lloyd, 1999; Rudolph et al., 2010a; Vincent et al., 1996). R-loops readily form behind transcribing RNA polymerases at sites of replication-transcription conflicts (Gowrishankar et al., 2013) and *in vitro* data using plasmid based assays conclude that RecG unwinds R-loops in a dose response manner, which removes the ability of R-loops to be used as a primer for plasmid replication (Fukuoh et al., 1997). By removing R-loops that impede replication fork progression, RecG may facilitate fork movement through highly transcribed areas, a feature also shared by UvrD helicase (Wollman J. et al., 2023). Loss of RecG's R-loop unwinding ability may therefore allow increased R-loop persistence that can block replication forks and cause instability. These unrestrained fork collisions likely generate substrates for pathological re-initiation events mediated by R-loops when RecG is also missing (Dimude et al., 2015, 2018a; Rudolph et al., 2009). In addition to processing R-loops directly, several other facets of RecG's multifunctional nature may contribute to suppressing R-loop associated instability Dimude et al. (2015).

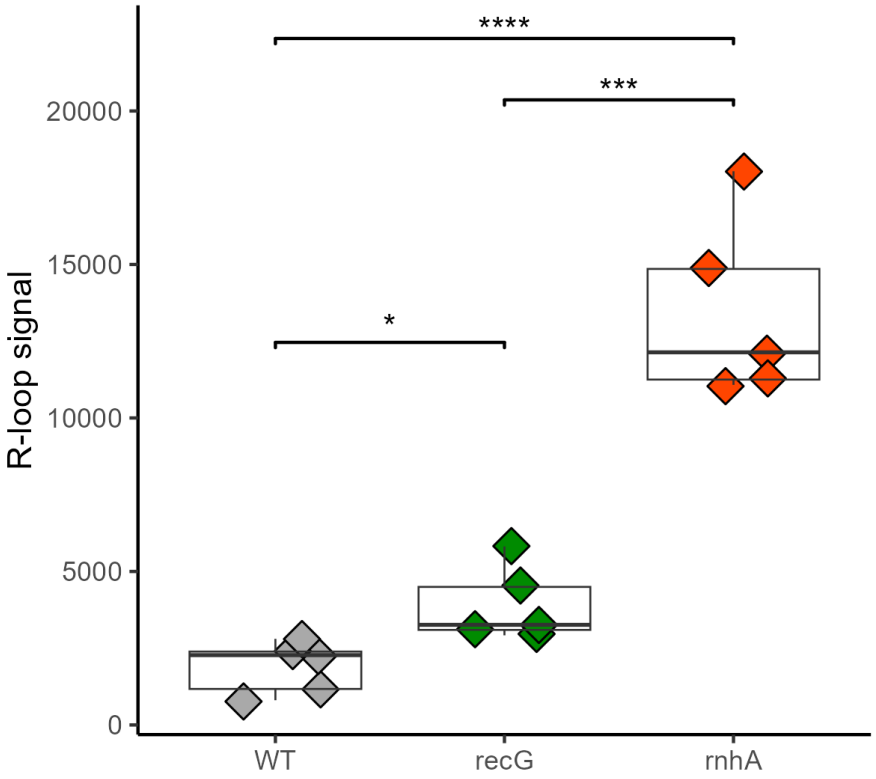

Figure 35: Results show statistically significant increase in R-loop levels in *recG* cells compared to WT (ANOVA p = 0.01). 5 experimental repeats were used to calculate the changes in R-loop signal in *recG*$^{-}$ mutants. *rnhA* mutants showed as the positive control for R-loop signal, and WT cells showed the lowest levels. Box plot overlay was used to visualise distribution of data points.

RecG's Holliday junction branch migration activity helps resolve recombination intermediates that can arise when replication forks collapse at R-loop roadblocks (McGlynn and Lloyd, 1999). Fork restart depends on RecG remodelling D-loops through regression or generating Holliday junctions for RuvABC resolution. Furthermore, RecG facilitates proper replication completion by converting 3' flap structures at fork fusion intermediates into 5' flaps not targeted by PriA restart pathways. Unrepaired 3' flaps may promote instability via R-loop formation and subsequent SDR . While modest, the increased R-loop levels observed in *recG* mutants likely compound with defects in processing other problematic DNA structures to promote genome instability when RecG helicase is absent.

rnhA mutants showing over-replication of termination area, as do *recG* mutants, but the previously hypothesized initiation starting from R-loops in *recG* cells is not the case, and genetics data show that the over-replication observed is likely due to 3' flap formation as a consequence of fork fusions, and not initiation from R-loops inside the termination area (Dimude et al., 2015; Maduike et al., 2014). There is no evidence that fork fusions have a direct impact on R-loop for-

mation, and the over-replication of the termination area seen in *rnhA* is likely down to global R-loop levels in these cells, and initiation taking place at an R-loop inside the termination area will be trapped by the fork trap and Tus-*ter* barriers, leading to the defined peak seen in marker frequency analysis.

One might hypothesize that in *recG* cells, RNase HI simply cannot gain access to the R-loop structure due to topological perturbations, or that different sized R-loops are better processed by either helicases or nucleases. Larger R-loop regions would indeed cause more topological changes what may require helicase activity for relaxation before RNase HI can properly perform its function. Elucidating precise mechanisms connecting RecG's diverse activities to R-loop suppression and replication fidelity remain an important ongoing goal and opens the question of why the cell would need helicase activity to unwind RNA-DNA hybrids, when nuclease activity should be sufficient and more of a permanent solution to RNA-DNA hybridization?

From the early *in vitro* data of RecG acting on R-loops, one might hypothesize that *recG* single mutants have comparable levels of R-loops from gDNA extracts as with *rnhA* mutants (Vincent et al., 1996). However, this is not what was observed from *recG* mutant gDNA extracts after being grown to an OD of 0.4, seen from Figure 35. More recently, dot blotting with S9.6 antibody did not show elevated levels in *recG* single mutants (Raghunathan et al., 2019). With these drastically different observations and contradictions between *in vitro* and *in vivo* data, a more statistically driven *in vivo* approach was chosen here to determine the true role RecG has on R-loop accumulation *in vivo*, and this data shows only a modest increase in *recG* single mutants, however this assay does not measure mechanism so only so much insight can be provided from *recG* single mutants.

RNase HI and RecG have Synergistic effect on regulating R-loops *in vivo*.

To test the involvement of RecG on removal of R-loops *in vivo*, I wanted to investigate the R-loop accumulation in *recG rnhA* double mutants. The lethality of this double mutant has been suggested to result from R-loop toxicity (Hong et al., 1995; Kogoma, 1997), providing accurate measurements of normalized gDNA should provide a piece to the puzzle of the role of RecG in removal of R-loops and R-loop dependent toxicity *in vivo*.

We utilized a plasmid system to modulate RecG in an *E. coli* this double mutant. Controlling RecG production enabled analysis of R-loop levels when RecG is present versus depleted. Quantification of S9.6 immunoblots, seen in Figure 36, demonstrated a significant R-loop increase in the cultures supplemented with 0.2% glucose, opposed to 0.2% arabinose, indicating RecG normally counteracts R-loop accumulation. This reveals an essential requirement for RecG in constantly resolving R-loops arising throughout the chromosome. This increase in R-loop levels can be the source of over-replication through aberrant initiation via PriABC pathway (Dimude et al., 2015, 2018a; Kogoma, 1997; Midgley-Smith et al., 2018).

This data show an correlation of R-loop levels and in the absence of both RecG and RNase HI *in vivo*, meaning the unresolved R-loops accumulate, leading to global genomic instability and toxicity, in line with the idea of aberrant SDR (Hong et al., 1995; Kogoma, 1997) and that over expression of RecG from *pDM104* reduces this phenotype (Dimude et al., 2015). However, this single time point data set does not showcase the dynamics of R-loop accumulation over time when RecG is not present, so what does the R-loop landscape look like in a *recG rnhA* double mutant over time?

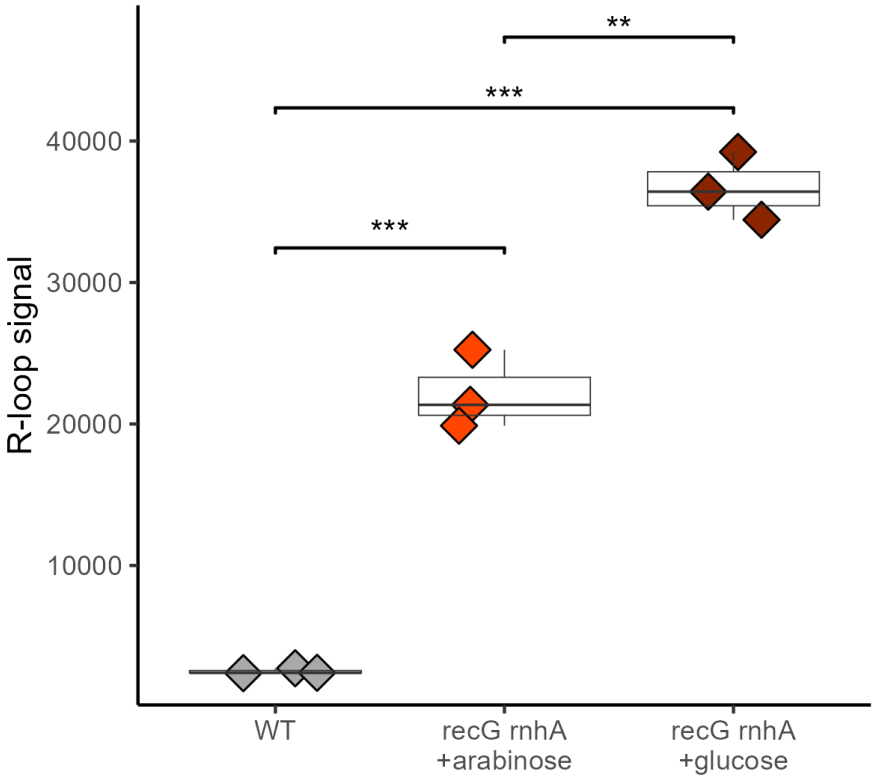

Figure 36: Quantification of R-loop signal in WT cells, *rnhA recG* double mutants supplemented with ampicillin and arabinose (*recG induced*), or ampicillin glucose (*recG* suppressed). 3 independent experiments were used to calculate the difference between the conditions. Results show significant increase in double mutants when supplemented with glucose vs arabinose, showing without RecG present the levels in gDNA extracts are elevated.

To test the hypothesis that excess R-loop accumulation is responsible for the lethality of *rnhA recG* double mutants when RecG is not expressed, I wanted to observe the level of R-loop accumulation overtime in this background when RecG is not expressed anymore. To do this a culture was grown to OD0.4 in Mu medium, cells were harvested and the media swapped from being supplemented with 0.2% arabinose to 0.2% glucose. Samples were taken every hour for 5 hours (T0-T5), gDNA extracted a dot blotted with the S9.6 antibody (Crossley et al., 2019; Raghunathan et al., 2019; Vlachos-Breton and Drolet, 2022).

If excess R-loop build up is responsible for the lethality of *rnhA recG* mutants then we would expect the signal to increase overtime. This is exactly what I observed. The time series dot blot showed that the level of R-loops steadily increases between T0-T2 and then drastically increases until T4, with a final decrease in signal at T5. This is in line with the hypothesis and shows that R-loop accumulation indeed significantly increases over time when RecG is absent.

The time course shows R-loop signal accumulates over time without RecG, eventually leading to the filamentation of cells and cell death, as previous work has demonstrated occurs in RNase HI-deficient mutants. There is a clear drop in signal between T4 and T5, indicating the R-loop accumulation becomes most toxic and induced lethality at some point during the experiment (Figure 37). Contrast enhancement reveals the T0 and T1 samples have barely detectable signal. The time series blot confirms low levels of R-loops between T0-2, before a drastic increase between T3-4 when RecG expression is inhibited. While the T4-5 signals appear identical, this is an artefact of contrast enhancement to visualize the earlier time points. In reality, the signal drops roughly 50% between T4-5. Figure 36 shows that R-loops do indeed increase drastically when RecG is not present as previously reported, linking the lethality of this double mutant to R-loop accumulation (Dimude et al., 2015; Gowrishankar et al., 2013; Kogoma, 1997).

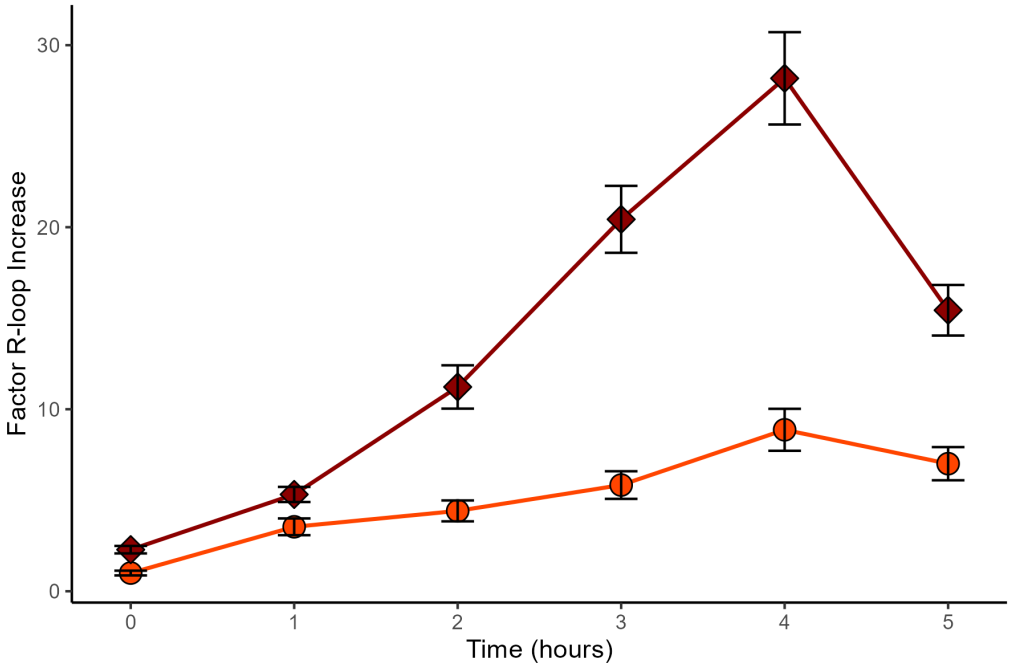

Figure 37: Time series data from *rnhA recG* double mutants supplemented with either arabinose or glucose, once in exponential phase to observe the change in R-loop accumulation over time without *recG* expression from pDIM104 plasmid. 4 experimental repeats were used to calculate error bars in the signal at each timepoint. Results show significant elevation at all timepoints, with a peak at T4, followed by a sharp drop in level at T5. R-loop level over time stayed mostly flat in the arabinose cultures.

The observed reduction in R-loop levels between the 4 hour and 5 hour timepoints likely does not solely reflect exceeding a toxicity threshold that causes population collapse and cell death. A more plausible explanation is that residual RecG protein present during the early timepoints continues to process R-loops for some time after switching the cells to glucose media.

Only once this residual RecG is exhausted do R-loop levels begin to rise dramatically, peaking around 4 hours as seen on the blot (Figure 37). The subsequent decline in signal at 5 hours could be partially explained by declining quantities of intact genomic DNA as cell viability starts to drop due to R-loop associated toxicity. With fewer living cells present after 4 hours without RecG, less overall DNA is being extracted, reducing the R-loop amount measured.

The decrease in R-loops between 4 and 5 hours is not solely due to exceeding a defined toxicity threshold. It may be partially explained by declining cell viability decreasing the DNA yield, combined with some remaining RecG activity in the initial timeframe temporarily suppressing accumulation. Accounting for residual protein effects in depletion experiments aids interpretation of results and kinetics underlying rapid R-loop escalation and toxicity in cells lacking RecG helicase activity.

Cell Death in *rnhA recG* Double Mutants is Detectable after two growth phases

The time course data revealed that R-loop levels decreased after approximately 4 hours of accumulation. This observation raised the question of why R-loop levels would drop instead of continuing to increase linearly over time? One hypothesis proposed was that 4 hours represents a critical R-loop toxicity threshold which cells can tolerate before population-wide collapse and cell death occurs. To explicitly test this model, cell viability was measured at defined timepoints with and without RecG expression.

To examine the effect of RecG depletion on cell viability, cultures were first grown in arabinose to an A_{600} of 0.4 to permit RecG expression. The culture was then split into two conditions - one supplemented with 0.2% glucose to inhibit RecG, the other supplemented with 0.2% arabinose to induce RecG expression. After reaching an initial OD of 0.4 in around 90 minutes, the first cell viability measurements showed only a minor reduction upon RecG withdrawal. The cultures were then diluted 1/100 in 11mL Mu medium under the same arabinose or glucose conditions and grown again to OD0.4 over another 90 minutes. At this point, the RecG-deficient culture exhibited a considerable loss of viability compared to the arabinose control, with no colonies formed even at the highest dilution when spotted on ampicillin arabinose plates that would induce RecG expression in viable cells. The depleted culture also took 30 minutes longer to reach the OD target. While less pronounced, the arabinose culture spotted at lower dilutions also appears less healthy than the initial culture, indicating a partial effect of the extended growth even with RecG present. Together, these results confirm RecG is required to maintain cell viability through successive growth cycles.

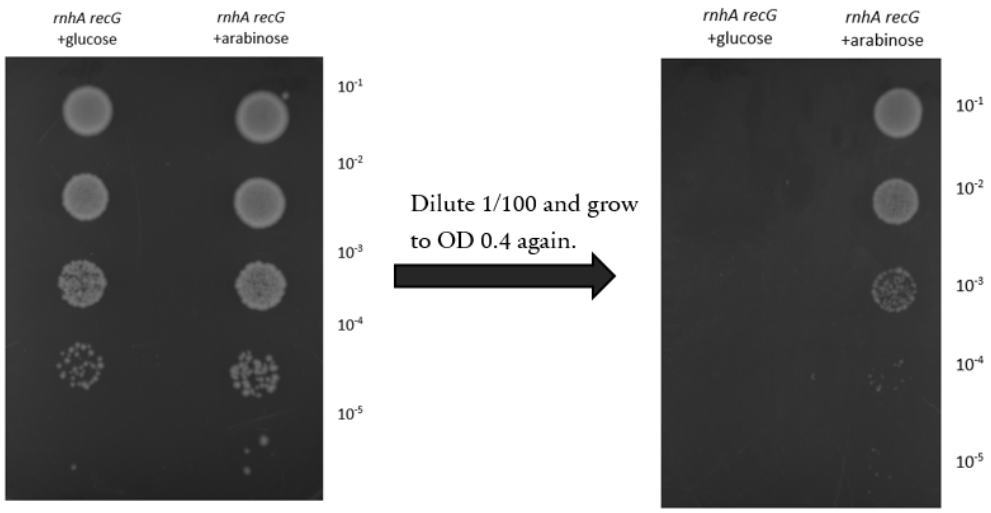

Figure 38: even though spot dilutions were carried out on Amp arabinose plates, meaning RecG expression could continue, there were still no viable colonies present for cultures with glucose supplementation showing that these cultures are not recoverable after the threshold accumulation of R-loops has been reached, indeed indicating onset of population wide cell death in these cells after ~4 hours of growth.

The results show that the culture with glucose indeed has less viable cells overall compared to the arabinose culture. Colonies cannot directly be counted between 10e-1 and 10e-3, however more fragmented spots are seen for the glucose+ condition. Indeed, this validates that without the expression of *recG* there are less viable cells inside a culture, even if they are then allowed to grow in favourable conditions (Figure 38). It should be mentioned that when the same serial dilution was performed on ampicillin glucose plates, there was no growth at all even at the highest dilution, further strengthening the idea that RecG is vital for the viability of *rnhA* mutants, where toxicity can be caused through excess R-loop accumulation. The dot blotting indeed showed accumulation of R-loops over time (Figure 38). This data adds to this and provides an explanation for the drop in R-loop signal after the 4 hour timepoint. Cell viability is decreasing and overall DNA concentration will also be declining, meaning no nascent R-loop synthesis which would be picked up by the assay as a decline in signal.

These results corroborate the model that unrestrained R-loop accumulation overwhelms cells during the second growth phase without RecG, precipitating widespread cell death. Measuring cell viability at specific timepoints during RecG withdrawal revealed substantial cell death initiated after 3-4 hours, closely matching the R-loop accumulation from the time course data. Taken together, these findings reinforce the hypothesis that uncontrolled R-loop accumulation triggers lethal toxicity in cells lacking RecG and RNase HI.

Taken together, these data show R-loops accumulate overtime in double mutants that lead to

pathological consequences from lethal R-loop accumulation over time. The genomic instability seen in these double mutants will be driven by aberrant replication at R-loop sites and dysregulation of DNA topology (Brochu et al., 2020; Drolet and Brochu, 2019), but also through flap structure accumulation that would normally be processed by RecG helicase (Dimude et al., 2015; Midgley-Smith et al., 2018; Rudolph et al., 2013).

UvrD accessory helicase and R-loops

I next examined whether UvrD helicase impacts R-loop homeostasis given its role in resolving conflicts between replication and transcription. A recent study showed that PcrA, the Gram-positive homolog of UvrD, can unwind R-loops efficiently *in vitro* via its 3' - 5' helicase activity when translocating along the DNA strand (Urrutia-Irazabal et al., 2021). They also found that blocking PcrA activity in Bacillus subtilis cells, either by overexpressing a dominant-negative mutant or preventing PcrA-RNA polymerase association, led to significantly increased R-loop levels as assessed by dot blotting with the S9.6 antibody.

In my experiments, deletion of *uvrD* in E. coli led to a moderate increase in R-loop levels, aligning with this prior work demonstrating a role for PcrA/UvrD helicases in suppressing R-loops *in vivo*. Interestingly, deletion of *recG* resulted in an even greater increase in R-loop levels compared to the uvrD single mutant. This contrasts with previous findings from Raghunathan and colleagues who reported no elevation of R-loops in a *recG* single mutant using dot blotting with the S9.6 antibody (Raghunathan et al., 2019). However, it is important to note that my conclusions are based on multiple replicates and carefully controlled experiments, lending confidence to the observed increase in R-loops in the absence of RecG.

Strikingly, combining *uvrD* and *recG* deletions in a double mutant led to a dramatic synergistic increase in R-loops that was significantly greater than the additive effects of the single mutants. The R-loop levels in the double mutant were comparable to the ΔrnhA positive control (Figure 39). This synergistic effect reveals that robust suppression of R-loops in E. coli requires the complementary activities of both UvrD and RecG.

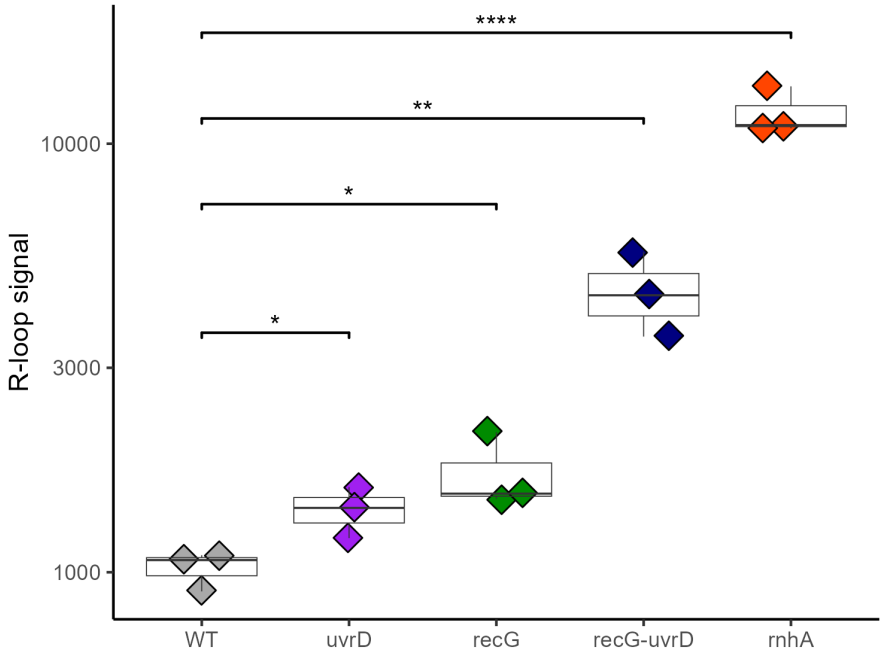

Figure 39: Effect of deleting *uvrD* and combining this with a *recG* deletion showing significant increases in R-loop accumulation in gDNA extracts. 3 experimental repeats were used to calculate difference between the strains, grown to a defined optical density. *uvrD* and *recG* single mutants showed comparable significance in the increase of R-loops, with *recG uvrD* double mutants showing a synergistic effect and significant elevation of R-loops present in gDNA extracts.

The substantial R-loop accumulation in the *recG* single mutant suggests that RecG plays a more prominent role in R-loop suppression than previously appreciated based on earlier results (Raghunathan et al., 2019). The synergistic effect of losing both UvrD and RecG could result from unrestrained replication-transcription collisions when these factors are missing, however this assay does not give mechanistic resolution, but loss of UvrD may allow conflicts that stall replication forks at R-loops and with RecG also absent, the RNA:DNA hybrids cannot be unwound efficiently, leading to massive accumulation. Earlier work by Boubakri and co-workers demonstrated that UvrD collaborates with DinG and Rep helicases to facilitate replication through highly transcribed regions, such as inverted rRNA operons (Boubakri et al., 2010) and Hawkins also recently proposed a model where UvrD displaces RNA polymerases to assist replication, preventing conflicts (Hawkins et al., 2019). My data also align with these results indicating that UvrD plays a role in suppressing R-loops during rapid growth when replication-transcription conflicts are prevalent.

The role of the replication fork trap

We next wanted to ask the question, is there a role for termination in R-loop formation? If there is a role for termination and fork fusions for generating R-loops, then deleting *tus*, a unique terminator protein, should give insights into the interplay between replication fork collision during termination, and R-loop formation. With replication fork dynamics known to have roles in regulating R-loop formation, identifying any perturbations in R-loops without the RFT would give more insight into the purpose of genomic architectures which organise the chromosome during replication [Maffia et al. (2020); Kumar and Remus (2023); Brochu2023].

R-loops readily accumulate when RNA-DNA hybrids are not adequately resolved, potentially blocking replication forks and eliciting DNA damage, as well as being sites of initiation themselves. I examined R-loop levels in cells lacking the fork-blocking Tus protein, which fails to properly arrest replication forks approaching from the permissive direction at *ter* sites. Although Tus shows no direct interactions with R-loop processing factors, I wanted to observe the effect of Tus on global R-loop levels. Intriguingly, R-loop levels were significantly increased compared to wild-type, although not as high as the *ΔrnhA* positive control, and inactivating the replication fork trap, therefore, was shown to elevate R-loop signal *in vivo* (Figure 40).

The increased R-loop levels observed in *tus* mutants raise intriguing questions about the interplay between replication fork dynamics and R-loop regulation. While Tus is not directly involved in R-loop metabolism, the disruption of fork fusion at *ter* sites appears to perturb R-loop homeostasis more substantially than inactivating individual R-loop processing enzymes like RecG or UvrD (Hawkins et al., 2019; Parekh et al., 2023). This suggests that proper fork convergence globally influences processes linked to R-loop generation across the chromosome.

One possibility is that loss of fork trapping at *ter* alters replication timing or origin activity in a way that promotes RNA-DNA hybrid formation. With replication forks no longer arrested by Tus, termination zones likely shift or become more diffuse across a wider region. This could expose sections of the chromosome to aberrant levels of supercoiling or transcriptional activity as they replicate at unusual times, allowing more R-loops to accumulate. Analyzing origin firing patterns and replication fork progression rates in *tus* mutants could lend insight into these potential genome-wide effects.

Additionally, widespread fork trap abolition may directly increase replication-transcription collisions outside of *ter* zones, providing more opportunities for R-loops to arise spontaneously

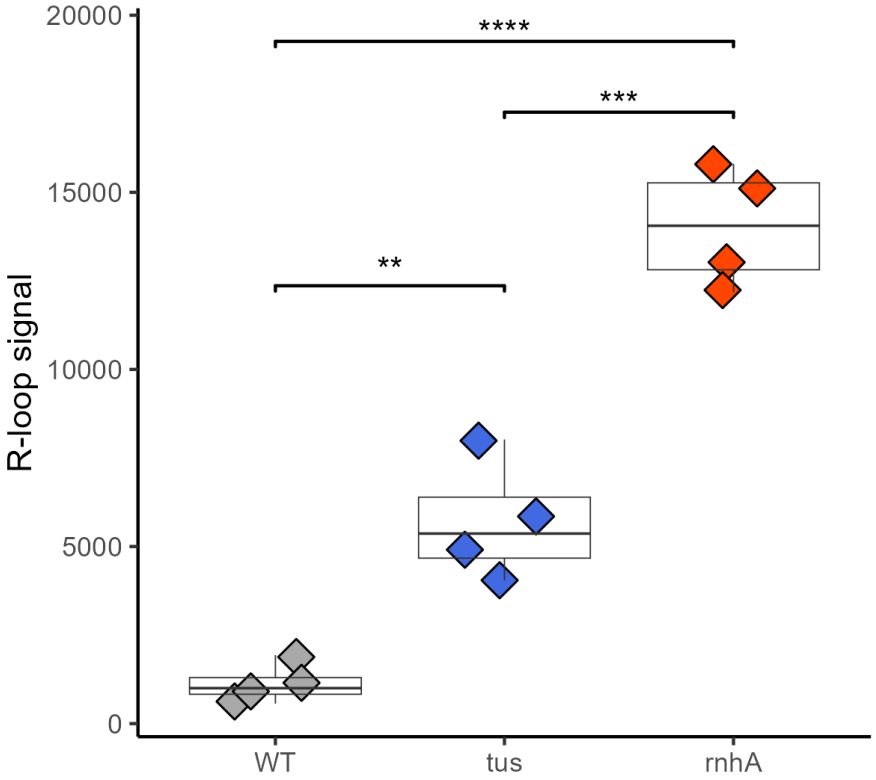

Figure 40: Quantification of *tus* mutant dot blots from gDNA extracts. 4 experimental repeats were used to calculate significance of the increase in R-loop level. Consistently, *tus* gDNA extracts showed higher levels of R-loops compared to WT gDNA. ANOVA p = 0.002 for *tus* data set.

when replisomes displace RNA polymerases. This hypothesis merits exploration by mapping R-loop distribution in *tus* cells to see if hotspots coincide with highly expressed genes. Overall, elucidating why loss of Tus elicits substantial R-loop escalation promises to uncover intriguing connections between chromosome organization, replication regulation, and RNA metabolism.

R-loop and G4 predictions in *E. coli*

While direct biochemical detection of R-loops and G-quadruplexes (G4s) across the entire *E. coli* chromosome was beyond the scope of this project, bioinformatics tools utilizing machine learning allow us to map genomic sites where these structures are predicted to form. Both R-loops and G4s preferentially arise in certain DNA sequences and contexts, so visualizing their genomic distributions can reveal insights into their potential functional roles. By utilizing predictive algorithms for R-loops (R-looptracker) and G4s (G4Boost), I aimed to observe where these structure forming sequences reside at chromosomal locations. This bioinformatics mapping of putative R-loop and G4 hotspots guides future inquiries into their functions and how they affect DNA replication.

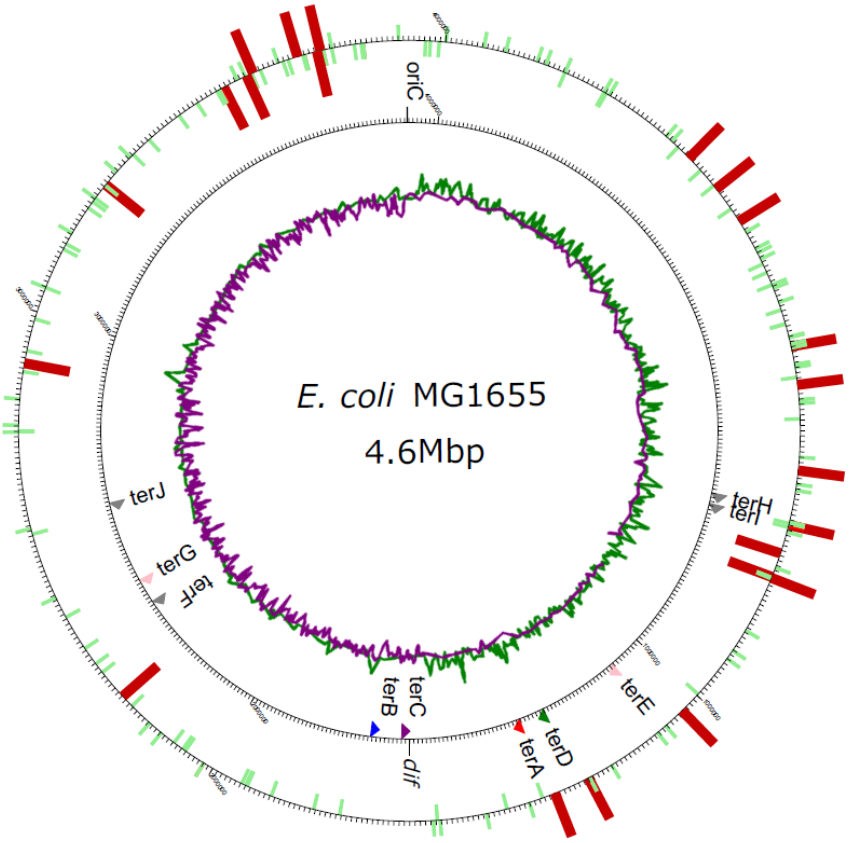

Figure 41: R-loop tracker was utilized as web tool, the raw csv was cleaned in custom R script and loaded into the custom python script utilising GenomeDiagram to map positions to the MG1655 genome. G4Boost was performed in the command line using command (*python G4Boost_v4.py –fasta MG1655.fasta –maxloop 10 –minloop 1 –maxG 4 –minG 3 –loops 4*). Blocks in red are large R-loop forming sequences ranging between 100 - 300bp long, and the green sites are G-quadruplex forming sequences ranging between 20 - 40bp long. Green and purple line graph shows GC skew of the MG1655 genome, a useful too predicting origin and termination sites, showing the skew aligns with *oriC* and *dif* sites.

Using the G4Boost algorithm (Cagirici et al., 2022), I searched the *E. coli* MG1655 genome for

G-quadruplex (G4) motifs under somewhat strict parameters, requiring loops between 1-10 nucleotides and 3-4 guanines per G-tetrad, with a total of 4 G-tetrads per motif. This returned ~130 hits which allowed for clearer visualization compared to a less restrictive search, which easily found ~100,000 motifs and created too much noise for visualisation (Figure 41). I correlated these predicted G4 locations with R-loop predictions from R-looptracker (Brázda et al., 2021), finding some co-localization and clustering of G4s and R-loops, mainly proximal to the origin of replication *oriC*. Intriguingly, both G4 and R-loop motifs appeared less saturated in the termination area of replication compared to other chromosomal regions. From figure 41 you are able to see only 2 R-loop forming sequences inside the termination area, closer to *terA* and *terD*, away from the *dif* site and *terC*. When comparing this result to the *tus* mutant R-loop levels *in vivo*, we can see little evidence for R-loops originating in the termination area.

The lower density of these structures in the termination area could potentially be explained by the GC-skew inherent to bacterial chromosomes, with the leading strand enriched in Gs and the lagging strand enriched in Cs. As R-loops form on transiently single-stranded DNA, they preferentially form on the G-rich lagging template strand, while G4s of course require guanine-rich sequences (Parekh et al., 2023). Figure 41 shows that the sequence of the termination area shows little activity for G4s and R-loop forming sequences in the MG1655 genome and that regions flanking *oriC* are more prone to formation of these DNA structures.

When visualizing the genomic distribution of R-loops using R-looptracker, I noted an asymmetry between the left and right chromosomal replichores, with more R-loop sequences concentrated on the right side, near the *rrn* operons and beyond. Interestingly, no R-loop or G4 forming sequences were observed near *terB* and *terC*, in which the *tus* gene resides, also proximal to the *dif* site (Figure 41). Though bioinformatics alone cannot determine the biological purpose of these sequence asymmetries, or the relative scarcity of sequences in the termination area, observing them guides hypotheses for future experimental investigations and this data highlights that possibly R-loops are not concentrated in the termination area.

By combining G4 and R-loop prediction tools, I have mapped hotspots where these structures are predicted to form. The clustering proximal to *oriC* compared to the termination region is intriguing and warrants further study into how replication fork dynamics shape the distribution of these motifs. Moreover, performing similar genomic visualizations for *E. coli* mutants like *ΔrnhA* and *Δtus* with known R-loop defects could reveal insights into how disrupting fork fusion or R-loop processing impacts sequence composition over evolution. In summary, over-

laying predicted non-B DNA structures guides both speculation about their purpose and design of future experiments elucidating their functions.

Chapter 3 - Discussion

RecG RNase HI interpretation

Intriguing questions are emerging about how R-loops lead to genomic instability and toxicity, and how RecG targets these structures *in vivo*. Elucidating these mechanisms by mapping R-loop structures precisely within *ter* will provide key insights (Hong et al., 1995). Defining pathways spatially and temporally regulating RecG's activity during termination represents an important future goal, although the assays used here to measure the effect of RecG on R-loops is not location specific and can be attributed to a more global increase of R-loops extracted from *recG* gDNA. With RecG active in many other processes throughout the chromosome, it is easy to see why its many functions have called RecG a general guardian of the genome (Rudolph et al., 2010b).

Surprisingly, *recG* single mutants exhibited increased R-loop signal compared to wildtype cells, contrasting what was previously shown (Raghunathan et al., 2019). In this system, RNase H is present and able to mitigate any additional R-loops arising due to loss of RecG. In contrast, the significant R-loop elevation in *rnhA* single mutants demonstrates RNase H's efficacy in removing R-loops.

The lethality of *rnhA recG* double mutants without a covering RecG plasmid indicates RecG alone cannot fully compensate for R-loop accumulation long-term in the absence of RNase H. However, RNase H appears able to sufficiently resolve R-loops without RecG. Thus, while RecG and RNase H may cooperate to process R-loops, RNase H plays the dominant role, as cells lacking RecG do not accumulate high R-loop levels as long as RNase H is present. The relatively modest R-loop signal in *recG* mutants shows that there isn't much of a detriment to the health of the cell, at least in terms of R-loop toxicity, when RecG is absent.

The results here reveal a complex interplay between replication fork dynamics, chromosome structure, and R-loop processing that is critical for maintaining genomic integrity. Intriguingly, R-loop levels were consistently increased in tus mutants compared to wildtype, affirming the global stabilizing role of the replication fork trap. The uncontrolled R-loop buildup in *recG rnhA* double mutants leads to rapid lethality, indicating key cooperation between RecG and RNase HI (Figure 36 and Figure 37), even if they are likely processing different substrates (Dimude et al., 2015).

It would be interesting to measure R-loop levels directly in the termination area to see if the R-

loop increase in RecG cells stems from the termination area specifically, where there is known to be over-replication, of if this is a more global phenomenon, currently we have no evidence for R-loops concentrated in the termination area. The *recG*- causes 3' flap accumulation, then over-replication, whereas *rnhA*- causes more R-loops and initiation via that pathway. Therefore, multiple overlapping pathways are at play for the lethality seen in *recG rnhA* double mutants.

Synergy between UvrD and RecG in limiting R-loops

The significant R-loop elevation in *uvrD recG* double mutants, compared to low levels in the respective single mutants, indicates UvrD and RecG may cooperate to constrain R-loops *in vivo*. With chapter 2 showing the same synergy in the recombination system, highlights the involvement of both helicases in both reversion of the resistance cassette, and *in vivo* R-loop levels. The R-loop data, however, could be direct or indirect evidence for R-loop formation, and this data alone does not show that RecG and/or UvrD are unwinding R-loops *in vivo*, simply that the global levels of R-loops present in gDNA is increased in these double mutants, highlighting synergy.

With UvrD being a helicase with 3' - 5' translocase activity, and RecG being able to translocate in both direction, the lack of these functions in global DNA repair would facilitate intermediates that arise at stochastic DNA damaging event, every cell cycle. When both key R-loop removal mechanisms are absent, formation outpaces degradation, rationalizing the dramatic synergistic increases in stable RNA-DNA hybrids.

The lack of RecG and UvrD presence in cells could result in increased replication-transcription collisions and R-loop formation in highly transcribed areas. RecG is needed to act across the genome to facilitate DNA repair and regular RecA, whereas UvrD is needed to backtrack stalled RNAP during transcription. Both helicases are shown to unwind R-loop substrates *in vitro*, however *in vivo* data is lacking. These results show that, both helicase are needed to mitigate R-loop levels *in vivo*.

In wild-type cells, UvrD uses its translocase activity to induce backtracking of RNA polymerases stalled on the lagging strand template ahead of approaching replication forks (Atkinson et al., 2009; Wollman J. et al., 2023). Backtracking exposes the mRNA-DNA hybrid to enable RNase HI access. Unlike Rep, UvrD does not travel along DNA associated with the replisome. However UvrD and Rep restart most blocked replisomes once they have collided with RNAP during a transcription conflict event. However, without UvrD, more polymerases will fully arrest forks, necessitating recombination pathways that further elevate R-loops.

During termination, without both UvrD and RecG, fork fusion events may generate structures prone to R-loop accumulation which cannot be resolved, even in the presence of RNase HI. Accumulation of flap structures would results in over-replication (Dimude et al., 2015; Rudolph et al., 2013; Rudolph et al., 2019). RecG is implicated in suppressing this over-replication through flap processing, with UvrD showing affinity for these same flap structures. UvrD and RecG could cooperate in various overlapping ways to constrain R-loops arising both proximal to and distal from the normal fork trap locus. The over-replication in *recG* and *3' exo* cells which is mitigated by PriA mutants, suggest flap structures indeed cause over-initiation of the termination area, determined through marker frequency analysis.

The bioinformatics mapping of R-loops and G4s indeed seem to suggest a lack of R-loop forming sequences inside the termination area. While it may be tempting to suggest that R-loops are concentrated in the termination area in *recG uvrD* double mutants, without molecular mapping of these structures, we currently do not have evidence that R-loops are actively being formed inside the termination area in these mutants.

Additional biochemical studies probing UvrD and RecG interactions could also lend mechanistic insight into how they work together at the molecular level. Elucidating this cooperative interplay is key to fully understanding how accessory helicases constrain threats to stability arising during final stages of chromosome duplication.

With RecG acting on Holliday junctions to promote branch migration and remove RecA filaments from ssDNA, and UvrD's role in MMR and NER recognised, this shows multiple mechanisms of genomic instability in *recG uvrD* double mutants. The results here do not identify mechanism of how these enzymes synergistically work to increase recombination frequencies and R-loop build up, but highlights that there are multiple measurable markers that are elevated in *recG uvrD* mutants.

R-loops and the replication fork trap

The results from inactivating Tus show that R-loop levels were significantly higher in *tus* single mutants, illustrating an important role for the replication fork trap in global chromosome stability. We know that the function of Tus acts, by definition, inside the termination area. With this result of higher R-loop signal in cells lacking Tus, one might hypothesize that the R-loops are being generated in the termination area in *tus* mutants.

Whether R-loops specifically accumulate proximal to the termination region in *tus* mutants is beyond the scope here but merits future investigation. Probing R-loop distribution across the genome could reveal if *tus* mutants exhibit greater R-loop buildup near the normal fork trap locus when Tus is absent. The bioinformatics predictions of R-loop and G4 forming sequences showed relatively quiet activity of R-loops inside the the termination area. While this result is not experimental, it does rule out sequence specific cause for the presence of R-loops between *terB* and *terD*.

Are increased collisions between replisomes and RNA polymerases in *tus* mutants directly responsible for generating R-loops seen in *tus* gDNA extracts? The loss of Tus would theoretically enable replication forks to proceed beyond the boundaries of the fork trap, although most forks still fuse proximal to *dif* and *terC*. More likely, without Tus, there is an increased range of possible fusion sites across the Ter macrodomain without Tus present (Hill, 1992; Rudolph et al., 2010c, 2013). Whether this alteration in local replication dynamics is causing the increased R-loop levels due to more frequent replication-transcription conflicts, remains to be seen, however it seems unlikely to be the case give the high level of R-loops in *tus* mutants and the minimal flattening of the copy number results from marker frequency analysis (Rudolph et al., 2010c, 2010b, 2013).

Given the genomic stabilizing effect of the replication fork trap, high resolution mapping of R-loop positions would delineate whether increased R-loops in *tus* mutants are concentrated at the terminus or distributed more broadly. Discovering the principles governing R-loop dynamics in the context of the fork trap to further elucidate the mechanisms for increased R-loops in *tus* mutants and if they are localised to a certain region, or more broadly distributed.

Further experiments mapping R-loop distribution in *tus* mutants via DRIP-seq could delineate if R-loops accumulate uniformly across chromosomes or are focused around the terminus when Tus is absent. This would help distinguish between global effects from fork movements versus local buildup near dysregulated fusion sites. Overall, the increased R-loops imply Tus plays an important role in maintaining chromosome stability that extends beyond its characterized replication fork trap function.

R-loops appear intricately tied to replication fork dynamics and genomic stability. Cells lacking Tus terminator protein exhibit global R-loop elevation, affirming the fork trap's role in chromosome stability. Surprisingly, RecG loss minimally increases R-loops, likely because RNase H is present for effective removal, and likely due to an indirect effect of RecG on R-loops, as previ-

ous research showed the cause of over-replication is different between *recG* and *rnhA* mutants (Dimude et al., 2015).

These findings reveal an unanticipated role for Tus-*ter* barriers in minimizing R-loops during replication, adding to prior models where the fork trap can act to reduce transcription-replication conflicts and prevent over-replication, although these data do not propose that minimizing RTC is the sole role of the fork trap as previously reported (Goodall et al., 2021, 2023; Ivanova et al., 2015; Rudolph et al., 2019). Spatial restriction of fork fusion events to Ter macrodomains evidently also lowers the probability of RTCs that foster R-loop formation. This work elucidates an unforeseen connection between replication termination control via the fork trap and maintenance of R-loop homeostasis.

This is a surprising finding because there is no known interaction of Tus and R-loops, and as Tus does not have helicase or ATPase activity, it is unlikely Tus is directly affecting R-loop persistence. Instead, I hypothesize that RTF inactivation leads to topological dysregulation can become unstable and lead to favourable conditions for R-loop formation. Negative supercoiling is a known factor for allowing R-loops to form and shows the importance of the topoisomerase enzyme family. The Drolet group have shown increased R-loop levels in *topA* mutants and that combinations of *topA* and *rnhA* are lethal to cells. Also that over expressing other topoisomerase enzymes (IV) can alleviate this phenotype (Brochu et al., 2020; Drolet and Brochu, 2019), illustrating the importance of regulating DNA supercoiling in R-loop formation. Without any direct ability for Tus to reduce R-loop levels through unwinding or degrading, I hypothesize that the loss of Tus causes topological dysregulation that is more favorable for R-loop formation (Brochu et al., 2023). It would be interesting to test topoisomerase mutants in combination with accessory helicases and Tus protein to determine mechanism of *tus*-dependent R-loop formation, however due to time constraints this was beyond the scope of this project.

R-loop and G4 co-localisation from prediction analysis

I wanted to examine whether there are R-loop forming sequences inside the termination area in MG1655 as a way to gain insight into DNA structure formation in this region. Using bioinformatics prediction analysis, I found that there were only 2 R-loop forming sequences inside the termination area, compared to a greater density proximal to *oriC*. G4s also appeared to have a lower density inside the termination area based on the predictions. This initial analysis revealed some co-localization between putative R-loop and G4 forming sequences in the genome,

supporting the idea that these structures may cluster together in certain genomic regions. The rationale behind analyzing the termination area was to see if R-loops detected in *tus* mutant genomic DNA could originate inside this region where replication terminates. However, the prediction data showed few native R-loop and G4 forming sequences inside the termination area in MG1655. While this does not reflect realistic R-loop levels *in vivo*, it does show that there should be no sequence specific reason for finding increased R-loops in the termination area.

This relative paucity of R-loops and G4s in the termination area leads to interesting speculation about the evolution of this region. The fork fusion complex or "fork trap" between Tus protein and *ter* sites is thought to have evolved to ensure proper chromosome segregation and reduce aberrant initiation taking place from the fusion of forks [Rudolph et al. (2013); Dimude et al. (2016); Goodall et al. (2021); @ Goodall2023]. Perhaps this fork trap system co-evolved with a sequence composition that minimizes R-loops and G4s, as these structures would likely impede the fusion of converging replisomes. It would be fascinating to examine the termination regions of bacterial species lacking Tus-*ter* systems to see if their sequences are enriched in motifs promoting R-loops and G4s.

While this bioinformatics analysis provided an initial glimpse into potential genome-wide distributions of R-loops and G4s, more experiments are needed to pinpoint actual locations where these structures accumulate in the mutants examined in this project. The prediction data served as a starting point to gain insight into regions where R-loops and G4s may preferentially arise, setting the stage for future work to more finely map their locations across the *E. coli* genome.

The accumulation of R-loops in *tus* mutants raises a question about their specific location within the genome. Given Tus's known role as a terminator protein, it would be logical to hypothesize that these R-loops are accumulating more specifically in the termination area. This would suggest that Tus might have an additional function in preventing R-loop formation at termination sites, beyond its established role in halting replication forks.

However, this hypothesis is not supported by the results obtained from the Rlooptracker and G4boost software. These bioinformatics tools, designed to predict and analyze the formation of R-loops and G-quadruplexes respectively, do not indicate a sequence-specific explanation that would localize the R-loops within the termination area. This suggests that the increased accumulation of R-loops in *tus* mutants is not due to a lack of Tus at termination sites, but rather points to a more global role for Tus in preventing R-loop formation.

Broader Implications and Model Refinement

The R-loop quantification experiments provide broader insights that help refine models of R-loop processing and toxicity. A key finding is that RNase HI plays the dominant role in resolving R-loops, as recG mutants show minimal accumulation. This implies other pathways including RNase HI can uphold R-loop homeostasis without RecG, though gradual buildup eventually leads to toxicity in *rnhA recG* double mutants. The results also reveal cooperation between the replication fork trap and accessory helicases in constraining R-loops.

These findings refine the model by establishing RNase HI as the primary R-loop processing activity required for viability, with RecG playing an important but secondary role. The lethality of *rnhA recG* double mutants also underscores that gradual R-loop accumulation eventually exceeds compensation capacity without activities of RNase HI and RecG. Additionally, the global chromosome stabilizing effect of Tus revealed by increased R-loops in its absence expands the known functions of the replication fork trap into mitigating global R-loop accumulation.

Concluding remarks

In closing, this research has uncovered a previously unknown link between two important processes that promote bacterial chromosome stability: replication termination at the fork trap site and prevention of excessive R-loop structures. Our findings reveal that proper fork trapping, mediated by the Tus-*ter* complex, helps restrain R-loop accumulation in cells. Unchecked R-loops can cause problems like DNA breaks and rearrangements. Therefore, by facilitating fork trap termination, the Tus-*ter* system indirectly suppresses harmful R-loop levels and protects genome integrity. While prior work recognized the importance of the fork trap and R-loop regulation individually, we are the first to connect these two mechanisms. Ongoing studies will further explore the molecular basis of this intersection and how defects in fork trapping impact R-loop homeostasis and genome stability. Given that impaired replication termination promotes chromosome abnormalities, fully deciphering the intricate protein interactions governing Tus-*ter* fork barriers remains essential. This will provide foundational knowledge of how bacteria successfully manage DNA replication to accurately duplicate their genetic material. In summary, our research uncovers a key role for the Tus-*ter* fork trap in R-loop regulation and chromosome maintenance, bridging two fields that were previously thought to be separate.

General Discussion

Overview and Tus-*ter* role

Termination has received little attention compared to initiation and elongation (Dewar and Walter, 2017; Rudolph et al., 2019) and studying chromosomal systems which have evolved around this final stage of replication can give us molecular insights into how the final stages of termination are concluded, as well as their evolutionary drivers. Termination in prokaryotes is far from a simple process, requiring recombination proteins and helicases to successfully conclude (Midgley-Smith et al., 2018; Midgley-Smith et al., 2019) and the conservation of the fork trap architecture across related phyla show the importance of this system to be maintained in the bacterial chromosome (Goodall et al., 2021, 2023). While there is a low probability that the Tus-*ter* complexes participate in the actual fork fusion event (Duggin and Bell, 2009; Ivanova et al., 2015; Rudolph et al., 2013), the data presented in this thesis highlight the importance of the fork trap system in regulating processes which alter the integrity of the chromosome, through accumulation of aberrant ssDNA structures that can result in local recombination and R-loop formation.

The core findings presented in this thesis support a model where the replication fork trap provides a benefit to cells in maintaining local genome stability through restricting the location in which forks can freely fuse, once entered into the termination area. The high conservation of the fork trap architecture across various phylogenetic groups, including *E. coli*, *Shigella*, *Salmonella*, and *Klebsiella*, as illustrated in chapter 1, indicate an evolutionary benefit that persists through the phyla. This strict conservation, which is not confined to open reading frames, suggests a significant evolutionary importance in maintaining this architecture, however, the benefits of this conservation remain unclear, especially when growth rates are only moderately affected without a functional system. With the data generated as part of this PhD thesis, the potential explanations for this conservation may be in mitigating recombination and R-loop levels *in vivo*

which are needed for maintaining chromosomal stability.

Chapter 2 and the recombination rate data show that when examining where forks fuse at different locations, we can gain insights into local chromosomal stability, using reversion rate and colony formation as measurement proxies. The frequency data suggests that the Tus protein might offer some benefit in the termination area, which is not detectable from growth curve data alone, but the difference is significant at the molecular level. The increased levels in *tus*$^{-}$ mutants suggest a dual effect of Tus in its role in recombination. When forks are held up at Tus-*ter*, they may be susceptible to RecBCD processing and RecA-dependent recombination, whilst also playing a role in reducing recombination through stably blocking forks. My data show that, when forks are indeed held up at Tus-*ter* complexes in *oriZ*$^{+}$ *tus*$^{+}$ strains, there is a significant increase in recombination rates at *narU-kankanMX4*, which could be down to recombination split between more 'passive' Tus-dependent recombination from RecBCD at stalled forks, and more 'active' Tus-dependent recombination in which an earlier stalled fork collides with nascent forks coming from *oriZ* that will generate DSB as the polymerase runs off the DNA.

Chapter 3 and the dot blotting data clearly show significant increased in gDNA extracts from *tus* mutants. When Tus is deleted in a single origin strain, there is a modest increase in reversion rates before adding in extra origins. However, when an extra origin is added and forks are stalled at Tus barriers, we observe a higher increase. Interestingly, deleting Tus in the presence of an extra ectopic origin does not result in an additive effect as one might expect. Instead, levels drop back down to near wild-type (WT) levels with a single origin. This phenomenon can be explained by the fork fusion location being shifted into the left-hand replichore in the *tus*$^{-}$ *oriZ+* mutant. Both methodologies show increased instability markers in these mutants, highlighting multiple features of genomic instability which are mitigated by functional fork trap systems in *E. coli*.

Further examination of RecG's role in R-loop processing could provide additional insights into what structures accumulate to result in cell death. The fact that *tus* mutants also have significant R-loop levels in genomic DNA (gDNA) extracts points to another feature that the Tus-*ter* system and replication fork trap mitigate global R-loop levels. This could be due to topological dysregulation in *tus* mutants when replisomes move freely through the termination area.

Intriguingly, the helicases RecG and UvrD show synergy across two methods of measuring genomic instability: recombination rates and R-loop quantification of gDNA. This dual synergy suggests overlapping pathways in which both RecG and UvrD function to maintain stability, both

locally and globally. This data provides valuable insights into how helicases affect chromosome dynamics in *E. coli* and underscores the importance of processing DNA during replication, where a specific termination feature is likely present.

The conservation of the fork trap architecture across various phylogenetic groups, its role in recombination and genomic instability, and the synergistic function of helicases RecG and UvrD in maintaining genomic stability are key findings from this discussion. These findings not only help us understand the evolutionary importance of maintaining the fork trap architecture but also offer insights into the mechanisms underlying genomic instability, addressing one of the major research goals stated at the beginning of this thesis, to elucidate the function and benefit of the specialized fork trap architecture during the final stages of replication. Future research should continue to explore these areas, particularly the role of R-loops and helicases in maintaining genomic stability and how the interplay of a fork trap system helps mitigate consequences of fork fusions and the termination events.

Reversion Rates and Fork Fusion Consequences

The study of reversion rates in chromosomal replication presents a fascinating insight into the intricate mechanisms of cellular biology. One of the key findings is the role of fork fusion in increasing reversion rates. Fork fusion is a process where two replication forks, which are the areas where DNA is being replicated, merge. This fusion has been observed to cause a significant increase in reversion rates, which is a measure of how often a change in the DNA sequence is corrected or reversed.

Interestingly, the introduction of new origins of replication, which are the points where DNA replication begins, does not disrupt other areas of the chromosome unless a fork fusion event occurs. This suggests that any replichore, a unit of replication including an origin and termination point, with only one replisome, the molecular machine that carries out replication, will have a stable reversion rate comparable to that of a wild-type (WT) organism. The significant increase in reversion rates only occurs when forks fuse, a phenomenon that was observed at three locations on the chromosome.

The role of the protein Tus and its associated fork trap architecture in mitigating unwanted recombination also emerged as an important factor in this study. Recombination is the process by which DNA sequences are broken and rejoined in new combinations. It was observed that deleting Tus slightly increases reversion rates at the *narU* locus, a specific location on the chro-

mosome. This suggests that the fork trap architecture may be functioning to prevent unwanted recombination.

Further evidence supporting this hypothesis comes from recent data showing significant levels of R-loops in Tus mutants. R-loops are structures that form during transcription when the RNA molecule hybridizes with the DNA template, forming a stable RNA:DNA hybrid. These structures can interfere with normal DNA replication and repair processes, potentially leading to genomic instability. The increased reversion of the *kankanMX4* cassette, a genetic marker used in this study, in the Tus mutant background could be due to these increased R-loops.

These data point to the crucial role of the fork trap in controlling the location of fork fusions to mitigate any unwanted recombination between different areas of the chromosome. This mechanism appears to be a key factor in maintaining genomic stability during DNA replication. Considering there is a role for the replication fork trap, and therefore termination, in R-loop formation then we can broadly say that these data implicate one important feature of the replication fork trap in *E. coli* is to contain replication forks during fusion events to reduce acute genomic instability caused from fork fusions.

These results demonstrate increased recombination rates in the termination regions of both *oriZ* and *oriX* strains lacking Tus. Introducing an ectopic fork trap near recombination hotspots, like *yjhR* and *tldD* when ectopic origins are present, could potentially suppress this hyper-recombination phenotype, this would link the plasmid results from Hamilton and colleagues (Hamilton et al., 2023) to chromosome data, where they introduction of *ter* sites reduced RecBCD induced recombination and plasmid instability. Adding a fork trap complex to these sites would topological constrain replication forks as they converge, mimicking the native termination process. We could measure recombination frequencies near ectopic trap sites to test if fork traps are sufficient to reduce recombination regardless of genomic context. Observing decreased recombination would indicate the fork trap itself, not the specific terminator DNA sequences, prevents aberrant fork collisions that lead to damaging crossover events.

R-loops and Chromosome Dynamics

RNase HI and RecG mutants are known to be inviable, with proposed R-loop toxicity as the cause (Hong et al., 1995; Kogoma, 1997). RNase HI and UvrD mutants are also inviable, with increased R-loops seen even in *uvrD* single mutants (Urrutia-Irazabal et al., 2021). This work is the first to show a synergistic effect of *recG uvrD* double mutants, with R-loop accumulation quantified

through immunoblotting using the S9.6 antibody.

There are similarities between RecG and UvrD in terms of 3' ssDNA flap processing during termination. Both helicases translocate in the 3'-5' direction and are proposed to be involved in processing 3' flaps that occur during replication fork convergence at termination sites. The moderate increase in R-loops in *recG* single mutants does not indicate RecG is directly unwinding R-loops *in vivo*. Rather, it shows that cells lacking RecG helicase have more favorable conditions for R-loop formation, as also seen in *uvrD* mutants.

One possibility is that the lack of either helicase leads to changes in DNA topology that favour R-loop formation, rather than the helicases being directly involved in R-loop unwinding. The changes in DNA torsion in the absence of these helicases may mean RNase HI is unable to access R-loop structures and degrade them, despite still being present. The synergy between *recG* and *uvrD* double mutants in accumulating more R-loops, even with RNase HI present, suggests reduced RNase HI activity simply from lack of access to the structures due to topological issues.

The recent discovery that G-quadruplex (G4) structures form extensively in the *E. coli* genome and challenge stability provides a broader context for interpreting my results on RecG's roles (Parekh et al., 2023). While my data show RecG suppresses instability and gradual R-loop accumulation, Parekh et al. reveal RecG and related helicases DinG and RecQ also resolve G4 structures. This implies RecG acts at stalled replication forks to unwind various secondary structures like G4s and R-loops. Without RecG, these structures likely persist and increase mutations and recombination.

I propose RecG helps maintain genome integrity at difficult-to-replicate loci prone to collisions between replication and transcription, where structures like G4s and R-loops arise. Loss of RecG's activity allows fork progression problems at these sites, with RecG known to be localised to sites of fork stalling (Bianco, 2015). G4s may directly hinder fork movement or topoisomerases, and also co-localize with R-loops, further stabilizing them against RNase HI removal. Defining RecG's specificity and coordination with other enzymes at fragile genomic regions will provide key insights into pathways suppressing localized instability arising from replication-transcription conflicts.

DNA topology is a known factor in R-loop formation, with topoisomerase mutants also showing R-loop accumulation and chromosome segregation difficulties. The unexpected Tus dot blot results, where Tus interacts with *ter*-DNA but not directly with R-loops, also point to topological dysregulation in *tus* mutants that could favor R-loop accumulation. Overall, the data suggests

helicases RecG and UvrD cooperate to maintain proper DNA topology to suppress R-loop formation, with loss of both creating synergistic effects on R-loop accumulation even in the presence of RNase HI. Testing R-loops in *rnhA tus* double mutants, as well as *uvrD tus* would also provide insight into the role of the replication fork trap in R-loop maintenance, chromosome topology and genomic instability. The data presented in this thesis show that the fork trap architecture indeed has a role to play in maintaining genomic stability while fork fusions take place, and it seems that once the species has obtained this architecture, likely from plasmid based systems (Galli et al., 2019), then it is conserved in future generations and lineages, providing the cell with an additional system to deal with all the chaos which can result from fork fusions (Goodall et al., 2021, 2023).

While loss of Tus subtly impacts growth under optimal conditions, replication fork collisions in the termination area may exacerbate DNA damage in stressful environments. We could examine how DNA-damaging agents like UV and mitomycin C affect recombination rates and viability in *tus+* and *Δtus* strains. If fork traps become more critical for maintaining genome stability under DNA damage, we would expect *Δtus* strains to exhibit greater sensitivity along with heightened recombination. Investigating how DNA damage impacts termination phenotypes will reveal whether fork traps provide an important protective role when replication is disrupted.

Overall, these proposed experiments blending ectopic fork traps and DNA damage conditions would further test the model that Tus constrains homologous recombination by avoiding deleterious fork collisions. Observing if fork trap activity suppresses recombination regardless of context and becomes more critical under stress would demonstrate this anti-recombinogenic effect represents a key evolutionary function of replication fork traps.

My findings reveal that RecG depletion rapidly causes irreversible viability defects in *E. coli* likely stemming from uncontrolled R-loop accumulation. This poses the critical question of whether removing the inciting R-loops could restore genome stability after RecG loss, or if an irreversible tipping point is crossed. We could address this central issue by utilizing tunable plasmid systems to modulate other key factors influencing R-loop metabolism beyond RecG. For instance, expressing *rnhA*, the catalytic enzyme that directly degrades R-loops, from an inducible plasmid after RecG depletion may provide complementary resolution of pathological R-loops. If viability and morphology returns upon *rnhA* induction despite continued RecG deficiency, it would strongly indicate that direct removal of toxic R-loops alone is sufficient to reverse the genome instability cascading from RecG loss. This would demonstrate that while excessive R-loops trig-

ger inviability, timely resolution by RNase H can still restore genome homeostasis after initial RecG depletion.

Alternatively, ongoing inviability despite induced *rnhA* expression would imply that unchecked R-loop accumulation crosses an irreversible threshold that cannot be reversed solely by degrading the original R-loops. This would suggest additional mechanisms like extensive DNA damage or mutation accumulate as a consequence of initial runaway R-loop formation, creating instability that cannot be mitigated by simply removing the original R-loops after a critical point. Determining if direct R-loop degradation can compensate for RecG loss even after initial toxicity sets in will reveal pivotal insights into whether R-loops themselves are the root cause of lethality or if they instigate an irreversible cascade of genomic disruption.

We could also utilize plasmid systems to modulate topological regulators like TopoI that reduce negative supercoiling, which the drolet lab have shown clearly impacts R-loops levels (Brochu et al., 2023). Overexpressing TopoI after initial RecG depletion may suppress continued R-loop accumulation and toxicity by relaxing superhelical tension. If viability improves upon TopoI induction, it would confirm that RecG acts primarily indirectly by opposing R-loop favoring topology rather than directly resolving R-loops. This would demonstrate that the replication stress caused by RecG loss arises as an indirect effect of topological changes that promote pathological R-loop structures, which can be reversed by Topo I restoring superhelical equilibrium.

Additionally, plasmid expression of Tus may complement R-loop suppression in termination regions of *Δtus* strains where unchecked fork collisions likely disrupt replisome processivity. Restoring fork trap activity could prevent pathological R-loops by avoiding the transcriptional-replicational conflicts that produce them when forks collide. Observing R-loop suppression from Tus despite its lack of helicase/nuclease functions would conclusively demonstrate fork traps inhibit R-loops indirectly by sustaining processive replication.

In summary, utilizing RecG expression plasmid in *rnhA*$^{-}$ *recG*$^{-}$ double mutants provides a powerful approach to disentangle the complex interplay underlying pathological R-loop formation. Determining if complementary R-loop removal or topological modulation can compensate for RecG loss and rescue viability will reveal pivotal insights into the mechanisms, direct versus indirect, underlying RecG's role in preventing runaway R-loop mediated genome instability. More broadly, these functional experiments focused on reversing toxicity will significantly advance understanding of the intricate balance between replication, transcription and topology that maintains R-loop homeostasis, preventing the genome catastrophes that arise when this

equilibrium is disrupted.

Fusion model and link with eukaryotic replication

Fundamental differences in the architecture of replication forks between bacteria and eukaryotic cells necessitate distinct molecular pathways to facilitate faithful termination of DNA synthesis. As prior work has established in prokaryotes, termination mechanisms involve specialized helicases that migrate convergent forks, unwind the DNA duplex, and process joint molecule intermediates containing 3' flaps (Dimude et al., 2016; Midgley-Smith et al., 2018; Midgley-Smith et al., 2019; Rudolph et al., 2013). However, recent structural analyses reveal that in eukaryotes, the leading strand helicase encircles the DNA template, ultimately generating a 5' flap upon fork convergence. This divergence in fork architecture and termination intermediates underscores why bacterial termination pathways cannot be simplistically extrapolated to eukaryotic chromosomes. Instead, the distinct intermediates shaped by eukaryotic replisomes likely require specialized processing by 5' flap nucleases like FEN1 [Xu2018; Laverde2023].

Intriguingly, though the termination intermediates vary between domains, common regulatory themes emerge in managing replication fork dynamics. Just as prokaryotic termination helicases dissolve DNA-protein interactions at the fork, eukaryotic cells employ helicases such as Pif1 to remove replisomes after use (Malone et al., 2022) and FEN1 to degrade 5' flaps following fork collisions (Gloor et al., 2010; Orans et al., 2011). This process interlinks with ubiquitin-mediated degradation of replisome components, highlighting coordinated control in resetting replication machinery post-termination. Moving forward, elucidating the interplay between Pif1 remodeling at eukaryotic termination forks and dissolution of replisome architecture will provide key insights (Deegan et al., 2019). While bacterial studies have seeded initial termination models, emerging structural distinctions underscore the need to elucidate specialized eukaryotic pathways shaped by the unique architecture of chromosome duplication forks. Defining these termination mechanisms will be pivotal to preserving genomic fidelity and stability in eukaryotes.

Supplementary

S1 - Phylogenetic Tree Analysis:

Phylogentic analysis of all genomes involved in this study showed that strains cluster into groups which build on from previous works (Touchon et al., 2009). As reference genomes for each group were picked at random, creating a phylogenetic tree which shows similar clustering as in previous works was important in case there were any abnormalities in our analysis, i.e to be sure we haven't chosen strains which show a bias of similarity towards MG1655 (group A). This is shown by our inclusion of more diverse genomes of group D, such as with UMN026, and of group B2, such as with APEC078.

This analysis was performed on a single gene, Tus, as this story is one of the replication fork trap and I show that analysing the similarity of Tus protein across the groups is a good indicator of how similar strains are, give that some genomes have variation in the Tus-*ter* sequence (Figure 42). Although, overall there is nothing to suspect that any drastic binding changes are present.

It would be interesting to perform a phylogenetic analysis on the *ter* sites themselves, concatenated into one long gene-like string, to see if the clustering would be similar or perhaps even more in line with the previous group data. More intense phylogenetic trees have been built on a wider range of genes from all genomes in *E. coli*, and we use this analysis as our baseline, instead of performing the exact same analysis.

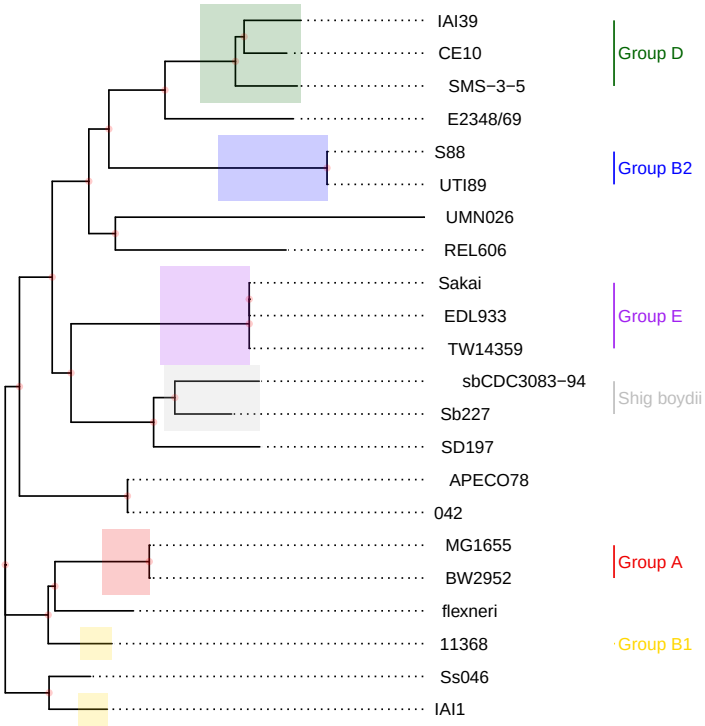

Figure 42: Tus phylogenetic tree shows group clustering using MSA alignment in R and rendered with ggtree Bioconductor packages, coloured boxes indicate group as previously identified with whole genome tree construction. Some outliers are present showing limitation of using only a single gene to perform phylogenetic tree construction.

S2 - Extended *ter* Analysis across *E. coli* Phylogroups:

To further illustrate the similarities between *ter* sites across phylogroups I created a python app which can calculate the edit distance of strings compared with MG1655, the reference wild type genome. Edit distance is defined as the minimal number of substitutions, deletions or insertions needed to turn one string into another, thus giving us a metric to work with for how similar *ter* sites are compared to MG1655.

The algorithm allows you to choose which between hamming distance and edit distance functions, however hamming distance does not take into account insertion or deletions and requires strings to be the same length. Although all *ter* sites analysed are 23 bases long, edit distance usually gives a clearer picture to how similar two strings are. For the purposes of our analysis, we can see little difference between hamming distance or edit distance of a comparison of MG1655 *terA* with *terA* of UMN026, for example. Where both algorithms return a score of 0, essentially meaning there is a no difference and both sequences are the same.

The nuance comes from edit distance being able to handle deletion or insertion possibilities when calculating its score, whereas hamming distance simply calculates substitutions needed. From this example, we can see that the similarities of the 23 base *ter* sites are relatively similar for *ter* sites *A-J*, not deviating more than 4. Whereas when we include the extended region which *ter* sites are located, we see much more variation in the scores essentially saying that the conservation of *ter* sites is higher than the surrounding chromosome, unless that *ter* site resides within an open reading frame, evident from the *terE-J* data, which is not surprising as there will always be high pressures to maintain genes. Using *terY* as a comparison of variation we would expect to see in the *ter* site sequences not within open reading frames, we can clearly see a difference in the primary RFT sites.

Figure 43 provides an insightful illustration of the conservation levels of various sites within open reading frames (ORFs). It is evident from the figure that sites located within ORFs exhibit a higher degree of conservation, which is indicative of the evolutionary pressure to maintain these critical regions.

The figure also provides a detailed view of the variation in the regions surrounding the termination sites (*ter*) in the *terABCD* sequence. Interestingly, while the surrounding regions show significant variation, the *ter* sites themselves are highly conserved. This suggests that the evolutionary pressure to maintain the *ter* sites is highly localized and does not extend to the wider

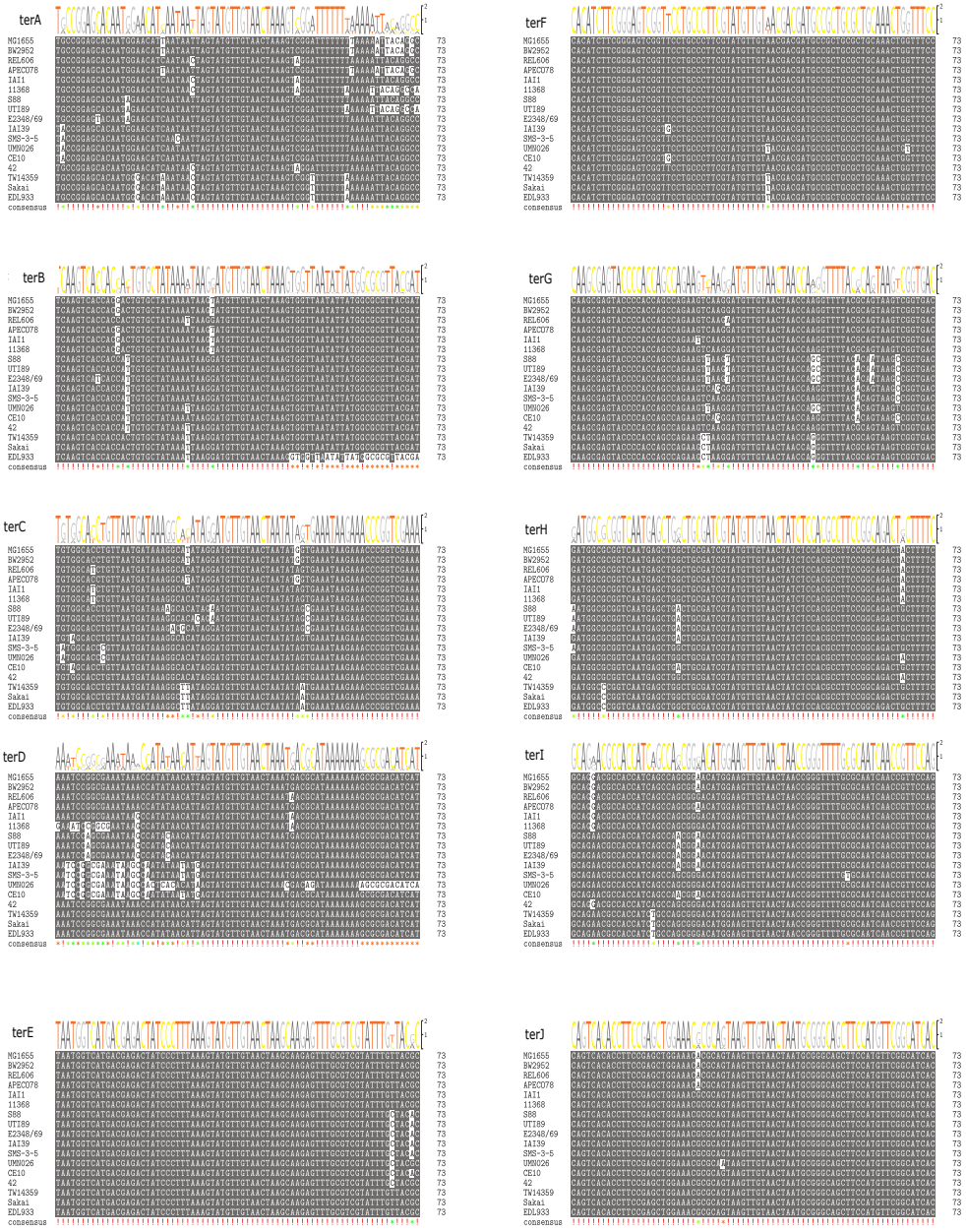

Figure 43: Larger window of alignment of *ter* sites across phylogenetic groups to see if there is a greater area of conservation of the local area, compared to the *ter* site itself. Those sites present in ORFs show higher local area conservation than those not present in ORFs.

region.

Furthermore, this pattern is particularly pronounced for inner *ter* sites. The conservation pressure appears to be confined strictly to the *ter* site itself, with no evident need for the wider region to be conserved. This observation underscores the importance of these specific sites in bacterial genetics and replication fork trap mechanisms in *E. coli*, and further highlights the specificity of evolutionary pressures in shaping bacterial genomes.

S3 - Dot blots

Figure 44 shows dot blot for *rnhA recG* double mutants grown in ampicillin and either supplemented with arabinose or glucose, grown to single timepoints to OD 0.4.

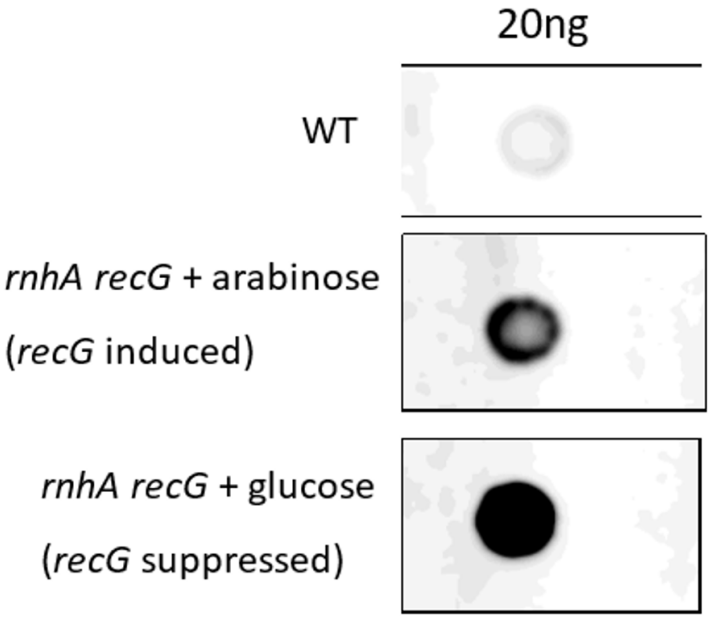

Figure 44: Results show statistically significant increase in R-loop levels in *rnhA recG* cells grown in glucose compared to arabinose. gDNA was normalized to 20ng and 2uL was spotted on hybond membrane. S9.6 was added 1/5000, secondary HRP-conjugated was added 1/10000. ECL HRP substrate was added for 5 minutes in dark and imaged on GBox for 30 seconds.

Figure 45 shows one example of the time series dot blots of *recG rnhA* supplemented with either arabinose or glucose grown in Mu (See methods for more details). two concentrations were made following extracting 20ng and 50ng. results show a sharp increase in signal over time in the glucose supplemented samples compared to arabinose, showing a clear role for RecG in toxic R-loop accumulation.

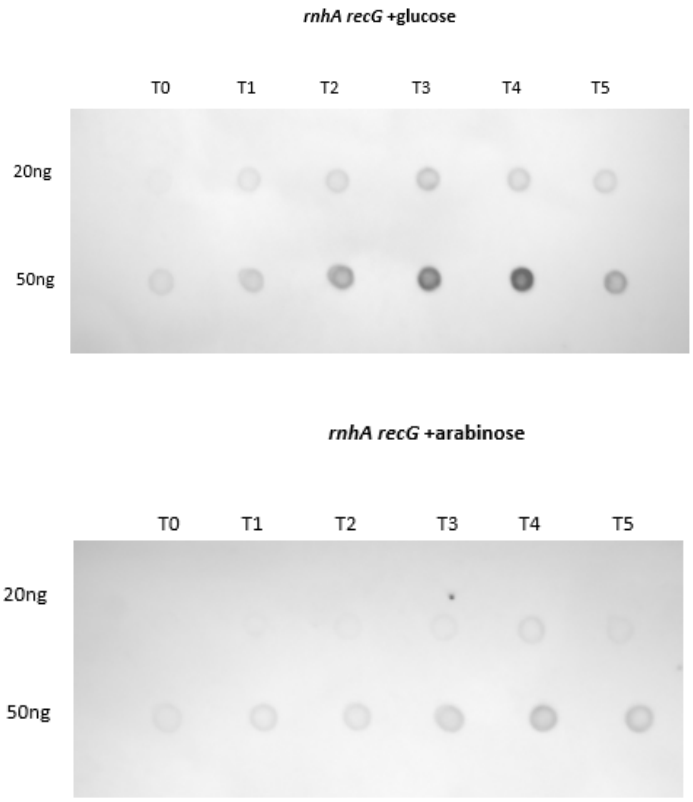

Figure 45: Dot blots were spotted 2uL using a P20 Gilson pipette to maintain a consistent dot blot to reduce the typical halo that can form. Results show statistically significant increase in R-loop levels in *recG* cells compared to WT ($p = 0.01$)

References

Abhyankar, M.M., Zzaman, S., and Bastia, D. (2003). Reconstitution of R6K DNA Replication in Vitro Using 22 Purified Proteins. Journal of Biological Chemistry *278*, 45476–45484.

Abram, K., Udaondo, Z., Bleker, C., Wanchai, V., Wassenaar, T.M., Robeson, M.S., and Ussery, D.W. (2021). Mash-based analyses of Escherichia coli genomes reveal 14 distinct phylogroups. Communications Biology *4*.

Al-Deib, A.A., Mahdi, A.A., and Lloyd, R.G. (1996). Modulation of recombination and DNA repair by the RecG and PriA helicases of Escherichia coli K-12. Journal of Bacteriology *178*, 6782–6789.

Amarh, V., White, M.A., and Leach, D.R.F. (2018). Dynamics of RecA-mediated repair of replication-dependent DNA breaks. Journal of Cell Biology *217*, 2299–2307.

Arthur, H.M., and Lloyd, R.G. (1980). Hyper-recombination in uvrD mutants of Escherichia coli K-12. MGG Molecular & General Genetics *180*, 185–191.

Asai, T., and Kogoma, T. (1994). D-loops and R-loops: Alternative mechanisms for the initiation of chromosome replication in Escherichia coli. Journal of Bacteriology *176*, 1807–1812.

Atkinson, J., Guy, C.P., Cadman, C.J., Moolenaar, G.F., Goosen, N., and McGlynn, P. (2009). Stimulation of UvrD helicase by UvrAB. Journal of Biological Chemistry *284*, 9612–9623.

Atlung, T., Clausen, E.S., and Hansen, F.G. (1985). Autoregulation of the dnaA gene of Escherichia coli K12. MGG Molecular & General Genetics *200*, 442–450.

Azeroglu, B., and Leach, D.R.F. (2017). RecG controls DNA amplification at double-strand breaks and arrested replication forks. FEBS Letters *591*, 1101–1113.

Azeroglu, B., Mawer, J.S.P., Cockram, C.A., White, M.A., Hasan, A.M.M., Filatenkova, M., and Leach, D.R.F. (2016). RecG Directs DNA Synthesis during Double-Strand Break Repair. PLoS Genetics *12*, 1–23.

Barre, F.X. (2007). FtsK and SpoIIIE: The tale of the conserved tails. Molecular Microbiology *66*, 1051–1055.

Bastia, D., Zzaman, S., Krings, G., Saxena, M., Peng, X., and Greenberg, M.M. (2008). Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. Proceedings of the National Academy of Sciences of the United States of America *105*, 12831–12836.

Berghuis, B.A., Dulin, D., Xu, Z.Q., Van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N.E., Depken, M., and Dekker, N.H. (2015). Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. Nature Chemical Biology *11*, 579–585.

Berghuis, B.A., Raducanu, V.S., Elshenawy, M.M., Jergic, S., Depken, M., Dixon, N.E., Hamdan, S.M., and Dekker, N.H. (2018). What is all this fuss about Tus? Comparison of recent findings from biophysical and biochemical experiments. Critical Reviews in Biochemistry and Molecular Biology *53*, 49–63.

Bianco, P.R. (2015). I came to a fork in the DNA and there was RecG. Progress in Biophysics and Molecular Biology *117*, 166–173.

Bianco, P.R. (2020). Dna helicase-ssb interactions critical to the regression and restart of stalled dna replication forks in Escherichia coli. Genes *11*, 19–23.

Bidnenko, V., Ehrlich, S.D., and Michel, B. (2002). Replication fork collapse at replication terminator sequences. EMBO Journal *21*, 3898–3907.

Bidnenko, V., Lestini, R., and Michel, B. (2006). The Escherichia coli UvrD helicase is essential

for Tus removal during recombination-dependent replication restart from Ter sites. Molecular Microbiology *62*, 382–396.

Bigot, S., Saleh, O.A., Lesterlin, C., Pages, C., El Karoui, M., Dennis, C., Grigoriev, M., Allemand, J.-F., Barre, F.-X., and Cornet, F. (2005). KOPS: DNA motifs that control E. coli chromosome segregation by orienting the FtsK translocase. The EMBO Journal *24*, 3770–3780.

Bird, R.E., Louarn, J., Martuscelli, J., and Caro, L. (1972). Origin and sequence of chromosome replication in escherichia coli. Journal of Molecular Biology *70*, 549–566.

Bishop, A.J.R., and Schiestl, R.H. (2002). Homologous recombination and its role in carcinogenesis. Journal of Biomedicine and Biotechnology *2002*, 75–85.

Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., and Sherratt, D.J. (1993). Two related recombinases are required for site-specific recombination at dif and cer in e. Coli K12. Cell *75*, 351–361.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453–1462.

Boe, L., Tolker-Nielsen, T., Eegholm, K.M., Spliid, H., and Vrang, A. (1994). Fluctuation analysis of mutations to nalidixic acid resistance in Escherichia coli. Journal of Bacteriology *176*, 2781–2787.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. Journal of Immunological Methods *89 1*, 123–130.

Boubakri, H., De Septenville, A.L., Viguera, E., Michel, B., H., B., Septenville A. L., de, E., V., and B., M. (2010). The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *29*, 145–157.

Brambati, A., Colosio, A., Zardoni, L., Galanti, L., and Liberi, G. (2015). Replication and transcription on a collision course: Eukaryotic regulation mechanisms and implications for DNA stability (Frontiers Media S.A.).

Brázda, V., Havlík, J., Kolomazník, J., Trenz, O., and Šťastný, J. (2021). R-loop tracker: Web access-based tool for R-loop detection and analysis in genomic DNA sequences. International Journal of Molecular Sciences *22*, 21–24.

Brewer, B.J. (1988). When polymerases collide: Replication and the transcriptional organization of the E. coli chromosome. Cell *53*, 679–686.

Briggs, G.S., Mahdi, A.A., Weller, G.R., Wen, Q., and Lloyd, R.G. (2003). Interplay between DNA replication, recombination and repair based on the structure of RecG helicase. Philosophical Transactions of the Royal Society B: Biological Sciences *359*, 49–59.

Briggs, G.S., Mahdi, A.A., Weller, G.R., Wen, Q., and Lloyd, R.G. (2004). Interplay between DNA replication, recombination and repair based on the structure of RecG helicase. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences *359 1441*, 49–59.

Brochu, J., Vlachos-Breton, É., Sutherland, S., Martel, M., and Drolet, M. (2018). Topoisomerases I and III inhibit R-loop formation to prevent unregulated replication in the chromosomal Ter region of Escherichia coli. PLoS Genetics *14*, 1–25.

Brochu, J., Breton, É., and Drolet, M. (2020). Supercoiling, R-Loops, Replication and the Functions of Bacterial Type 1A Topoisomerases.

Brochu, J., Vlachos-Breton, É., Irsenco, D., and Drolet, M. (2023). Characterization of a pathway of genomic instability induced by R-loops and its regulation by topoisomerases in E. coli.

Brüning, J.-G., Howard, J.L., and McGlynn, P. (2014). Accessory Replicative Helicases and the Replication of Protein-Bound DNA. Journal of Molecular Biology *426*, 3917–3928.

Cagirici, H.B., Budak, H., and Sen, T.Z. (2022). G4Boost: a machine learning-based tool for

quadruplex identification and stability prediction. BMC Bioinformatics *23*, 1–18.

Campbell, J.L., and Kleckner, N. (1990). E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. Cell *62*, 967–979.

Chodavarapu, S., and Kaguni, J.M. (2016). Replication Initiation in Bacteria. In Physiology & Behavior, pp.1–30.

Choudhary, R., Niska-Blakie, J., Adhil, M., Liberi, G., Achar, Y.J., Giannattasio, M., and Foiani, M. (2023). Sen1 and Rrm3 ensure permissive topological conditions for replication termination. Cell Reports *42*, 112747.

Clermont, O., Dixit, O.V.A., Vangchhia, B., Condamine, B., Dion, S., Bridier-Nahmias, A., Denamur, E., and Gordon, D. (2019). Characterization and rapid identification of phylogroup G in Escherichia coli, a lineage with high virulence and antibiotic resistance potential. Environmental Microbiology *21*, 3107–3117.

Cockram, C.A., Filatenkova, M., Danos, V., El Karoui, M., and Leach, D.R.F. (2015). Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo. Proceedings of the National Academy of Sciences of the United States of America *112*, E4735–E4742.

Costa, A., Hood, I.V., and Berger, J.M. (2013). Mechanisms for initiating cellular DNA replication. Annual Review of Biochemistry *82*, 25–54.

Courcelle, J. (2005). Recs preventing wrecks. *577*, 217–227.

Cox, M.M. (2001). RECOMBINATIONAL DNA REPAIR OF DAMAGED REPLICATION FORKS IN ESCHERICHIA COLI : Questions. Annual Review of Genetics *35*, 53–82.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000). The importance of repairing stalled replication forks. Nature *404*, 37–41.

Crosa, J.H., Luttropp, L.K., and Falkow, S. (1976). Mode of replication of the conjugative R plasmid RSF 1040 in Escherichia coli. Journal of Bacteriology *126*, 454–466.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Molecular Cell *73*, 398–411.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America *97*, 6640–6645.

De Graaff, J., Crosa, J.H., Heffron, F., and Falkow, S. (1978). Replication of the nonconjugative plasmid RSF1010 in Escherichia coli K-12. Journal of Bacteriology *134*, 1117–1122.

Deegan, T.D., Baxter, J., Ortiz Bazán, M.Á., Yeeles, J.T.P., and Labib, K.P.M. (2019). Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. Molecular Cell *74*, 231–244.e9.

Dennis, P.P., Ehrenberg, M., Fange, D., and Bremer, H. (2009). Varying Rate of RNA Chain Elongation during rrn Transcription in Escherichia coli. Journal of Bacteriology *191*, 3740–3746.

Dewar, J.M., and Walter, J.C. (2017). Mechanisms of DNA replication termination. Nature Reviews Molecular Cell Biology *18*, 507–516.

Dimude, J.U., Stockum, A., Midgley-Smith, S.L., Upton, A.L., Foster, H.A., Khan, A., Saunders, N.J., Retkute, R., and Rudolph, C.J. (2015). The consequences of replicating in the wrong orientation: Bacterial chromosome duplication without an active replication origin. mBio *6*, 1–13.

Dimude, J.U., Midgley-Smith, S.L., Stein, M., and Rudolph, C.J. (2016). Replication termination: Containing fork fusion-mediated pathologies in Escherichia coli. Genes *7*, 40.

Dimude, J.U., Stein, M., Andrzejewska, E.E., Khalifa, M.S., Gajdosova, A., Retkute, R., Skovgaard, O., and Rudolph, C.J. (2018b). Origins left, right, and centre: Increasing the number of initiation

sites in the Escherichia coli chromosome. Genes *9*.

Dimude, J.U., Midgley-Smith, S.L., and Rudolph, C.J. (2018a). Replication-transcription conflicts trigger extensive DNA degradation in Escherichia coli cells lacking RecBCD. DNA Repair.

Drolet, M., and Brochu, J. (2019). R-loop-dependent replication and genomic instability in bacteria. DNA Repair *84*, 102693.

Duggin, I.G., and Bell, S.D. (2009). Termination Structures in the Escherichia coli Chromosome Replication Fork Trap. Journal of Molecular Biology *387*, 532–539.

Duggin, I.G., Wake, R.G., Bell, S.D., and Hill, T.M. (2008). Micro Review The replication fork trap and termination of chromosome replication. *70*, 1323–1333.

Ede, C., Rudolph, C.J., Lehmann, S., Schürer, K.A., and Kramer, W. (2011). Budding yeast Mph1 promotes sister chromatid interactions by a mechanism involving strand invasion. DNA Repair *10*, 45–55.

Elshenawy, M.M., Jergic, S., Xu, Z.-Q., Sobhy, M.A., Takahashi, M., Oakley, A.J., Dixon, N.E., and Hamdan, S.M. (2015). Replisome speed determines the efficiency of the Tus–Ter replication termination barrier. Nature *525*, 394–398.

Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., Mironov, A., and Nudler, E. (2014). UvrD facilitates DNA repair by pulling RNA polymerase backwards. Nature *505*, 372–377.

Florés, M.J., Sanchez, N., and Michel, B. (2005). A fork-clearing role for UvrD. Molecular Microbiology *57*, 1664–1675.

Foster, P.L. (2000). Adaptive mutation: Implications for evolution. BioEssays *22*, 1067–1074.

Foster, P.L. (2006). Methods for Determining Spontaneous Mutation Rates. pp.195–213.

Fukuoh, A., Iwasaki, H., Ishioka, K., and Shinagawa, H. (1997). ATP-dependent resolution of R-loops at the ColE1 replication origin by Escherichia coli RecG protein, a Holliday junction-specific helicase. EMBO Journal *16*, 203–209.

Galli, E., Ferat, J.L., Desfontaines, J.M., Val, M.E., Skovgaard, O., Barre, F.X., and Possoz, C. (2019). Replication termination without a replication fork trap. Scientific Reports *9*, 1–11.

Gillet-Markowska, A., Louvel, G., and Fischer, G. (2015). bz-rates: A web tool to estimate mutation rates from fluctuation analysis. G3: Genes, Genomes, Genetics *5*, 2323–2327.

Gloor, J.W., Balakrishnan, L., and Bambara, R.A. (2010). Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading. Journal of Biological Chemistry *285*, 34922–34931.

Goodall, D.J., Jameson, K.H., Hawkins, M., and Rudolph, C.J. (2021). A Fork Trap in the Chromosomal Termination Area Is Highly Conserved across All Escherichia coli Phylogenetic Groups. International Journal of Molecular Sciences *22*, 7928.

Goodall, D.J., Warecka, D., Hawkins, M., and Rudolph, C.J. (2023). Interplay between chromosomal architecture and termination of DNA replication in bacteria. Frontiers in Microbiology *14*.

Gowrishankar, J., Krishna Leela, J., and Anupama, K. (2013). R-loops in bacterial transcription: Their causes and consequences. Transcription *4*, 153–157.

Hamilton, N.A., Wendel, B.M., Weber, E.A., Courcelle, C.T., and Courcelle, J. (2019). RecBCD, SbcCD and ExoI process a substrate created by convergent replisomes to complete DNA replication. Molecular Microbiology *111*, 1638–1651.

Hamilton, N.A., Jehru, A.E., Samples, W.N., Wendel, B.M., Mokhtari, P.D., Courcelle, C.T., and Courcelle, J. (2023). chi sequences switch the RecBCD helicase–nuclease complex from degradative to replicative modes during the completion of DNA replication. Journal of Biological Chemistry *299*, 103013.

Hamon, A., and Ycart, B. (2012). Statistics for the Luria-Delbrück distribution. Electronic Journal of Statistics *6*, 1251–1272.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*.

Hansen, F.G., and Atlung, T. (2018). The DnaA tale. Frontiers in Microbiology *9*, 1–19.

Hansen, F.G., Christensen, B.B., and Atlung, T. (2007). Sequence Characteristics Required for Cooperative Binding and Efficient in Vivo Titration of the Replication Initiator Protein DnaA in E. coli. Journal of Molecular Biology *367*, 942–952.

Hawkins, M., Malla, S., Blythe, M.J., Nieduszynski, C.A., and Allers, T. (2013). Accelerated growth in the absence of DNA replication origins. Nature *503*, 544–547.

Hawkins, M., Dimude, J.U., Howard, J.A.L., Smith, A.J., Dillingham, M.S., Savery, N.J., Rudolph, C.J., and Mcglynn, P. (2019). Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. Nucleic Acids Research *47*, 5100–5113.

Heller, R.C., and Marians, K.J. (2005). The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. Molecular Cell *17*, 733–743.

Heller, R.C., and Marians, K.J. (2006). Replication fork reactivation downstream of a blocked nascent leading strand. Nature *439*, 557–562.

Hendrickson, H., and Lawrence, J.G. (2007). Mutational bias suggests that replication termination occurs near the dif site, not at Ter sites. Molecular Microbiology *64*, 42–56.

Heo, A., and Park, W. (2015). Expression of the mexA gene requires the DNA helicase RecG in Pseudomonas aeruginosa PAO1. Journal of Microbiology and Biotechnology *25*, 492–495.

Hiasa, H., and Marians, K.J. (1992). Differential inhibition of the DNA translocation and DNA

unwinding activities of DNA helicases by the Escherichia coli Tus protein. Journal of Biological Chemistry *267*, 11379–11385.

Hidaka, M., Kobayashi, T., and Horiuchi, T. (1991). A newly identified DNA replication terminus site, TerE, on the Escherichia coli chromosome. Journal of Bacteriology *173*, 391–393.

Hill, T.M. (1992). Arrest of bacterial DNA replication. Annual Review of Microbiology *46*, 603–633.

Hill, T.M., Henson, J.M., and Kuempel, P.L. (1987). The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication. Proceedings of the National Academy of Sciences of the United States of America *84*, 1754–1758.

Hill, T.M., Kopp, B.J., and Kuempel, P.L. (1988). Termination of DNA replication in Escherichia coli requires a trans-acting factor. Journal of Bacteriology *170*, 662–668.

Hizume, K., and Araki, H. (2019). Replication fork pausing at protein barriers on chromosomes. FEBS Letters *593*, 1449–1458.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). Escherichia coli RecG and RecA proteins in R-loop formation. EMBO Journal *14*, 2385–2392.

Horiuchi, T., and Fujimura, Y. (1995). Recombinational rescue of the stalled DNA replication fork: A model based on analysis of an Escherichia coli strain with a chromosome region difficult to replicate. Journal of Bacteriology *177*, 783–791.

Horiuchi, T., Fujimura, Y., Nishitani, H., Kobayashi, T., and Hidaka, M. (1994). The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. Journal of Bacteriology *176*, 4656–4663.

Hyrien, O. (2000). Mechanisms and consequences of replication fork arrest. Biochimie *82*, 5–17.

Ivanova, D., Taylor, T., Smith, S.L., Dimude, J.U., Upton, A.L., Mehrjouy, M.M., Skovgaard, O., Sherratt, D.J., Retkute, R., and Rudolph, C.J. (2015). Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. Nucleic Acids Research *43*, 7865–7877.

Jain, K., Wood, E.A., and Cox, M.M. (2021b). The rarA gene as part of an expanded RecFOR recombination pathway: Negative epistasis and synthetic lethality with ruvB, recG, and recQ.

Jain, K., Wood, E.A., Romero, Z.J., and Cox, M.M. (2021a). RecA-independent recombination: Dependence on the Escherichia coli RarA protein. Molecular Microbiology *115*, 1122–1137.

Jameson, K.H., Rudolph, C.J., and Hawkins, M. (2021). Termination of DNA replication at Tus-ter barriers results in under-replication of template DNA.

Kaguni, J.M. (2006). DnaA: Controlling the initiation of bacterial DNA replication and more. Annual Review of Microbiology *60*, 351–371.

Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K. (1996). Structure of a replication-terminator protein complexed with DNA. Nature *383*, 598–603.

Katayama, T. (2001). Feedback controls restrain the initiation of Escherichia coli chromosomal replication. Molecular Microbiology *41*, 9–17.

Kitagawa, R., Mitsuki, H., Okazaki, T., and Ogawa, T. (1996). A novel DnaA protein-binding site at 94.7 min on the Escherichia coll chromosome. Molecular Microbiology *19*, 1137–1147.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiology and Molecular Biology Reviews : MMBR *61*, 212–238.

König, F., Schubert, T., and Längst, G. (2017). The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. PLoS ONE *12*, 1–13.

Kono, N., Arakawa, K., and Tomita, M. (2011). Comprehensive prediction of chromosome dimer resolution sites in bacterial genomes. BMC Genomics *12*, 19.

Kono, N., Arakawa, K., and Tomita, M. (2012). Validation of bacterial replication termination models using simulation of genomic mutations. PLoS ONE *7*.

Kovačič, L., Paulič, N., Leonardi, A., Hodnik, V., Anderluh, G., Podlesek, Z., Žgur-Bertok, D., Križaj, I., and Butala, M. (2013). Structural insight into LexA-RecA* interaction. Nucleic Acids Research *41*, 9901–9910.

Krabbe, M., Zabielski, J., Bernander, R., and Nordström, K. (1997). Inactivation of the replication-termination system affects the replication mode and causes unstable maintenance of plasmid R1. Molecular Microbiology *24*, 723–735.

Krašovec, R., Belavkin, R.V., Aston, J.A.D., Channon, A., Aston, E., Rash, B.M., Kadirvel, M., Forbes, S., and Knight, C.G. (2014). Mutation rate plasticity in rifampicin resistance depends on Escherichia coli cell-cell interactions. Nature Communications *5*, 1–8.

Kreuzer, K.N. (2005). Interplay between DNA replication and recombination in prokaryotes. Annual Review of Microbiology *59*, 43–67.

Kuemmerle, N.B., and Masker, W.E. (1980). Effect of the uvrD mutation on excision repair. Journal of Bacteriology *142*, 535–546.

Kuempel, P.L., Maglothin, P., and Prescott, D.M. (1973). Bidirectional termination of chromosome replication in Escherichia coli. MGG Molecular & General Genetics *125*, 1–8.

Kuempel, P.L., Duerr, S.A., and Seeley, N.R. (1977). Terminus region of the chromosome in Escherichia coli inhibits replication forks. Proceedings of the National Academy of Sciences of the United States of America *74*, 3927–3931.

Kuempel, P.L., Henson, J.M., Dircks, L.K., Tecklenburg, M.L., and Lim, D. (1991). Dif, a recA-independent recombination site in the terminus region of the chromosome of escherichia coli.

The New Biologist *3 8*, 799–811.

Kumar, C., and Remus, D. (2023). Looping out of control: R-loops in transcription-replication conflict. Chromosoma.

Kuzminov, A. (1995). Collapse and repair of replication forks in Escherichia coli. Molecular Microbiology *16*, 373–384.

Kuzminov, A. (2016). Chromosomal Replication Complexity: A Novel DNA Metrics and Genome Instability Factor. PLoS Genetics *12*, 1–20.

Kuzminov, A. (2018). When DNA topology turns deadly: Trends in Genetics *34*, 139–148.

Labib, K., and Hodgson, B. (2007). Replication fork barriers: Pausing for a break or stalling for time? EMBO Reports *8*, 346–353.

Lang, K.S., and Merrikh, H. (2018). The Clash of Macromolecular Titans: Replication-Transcription Conflicts in Bacteria. Annual Review of Microbiology *72*, 71–88.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., Merrikh, H., K. S., L., et al. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*.

Lea, D.E., and Coulson, C.A. (1949). The distribution of the numbers of mutants in bacterial populations. Journal of Genetics *49*, 264–285.

Lee, J.Y., and Yang, W. (2006). UvrD Helicase Unwinds DNA One Base Pair at a Time by a Two-Part Power Stroke. Cell *127*, 1349–1360.

Lemon, K.P., and Grossman, A.D. (2001). The extrusion-capture model for chromosome partitioning in bacteria. Genes and Development *15*, 2031–2041.

Lenhart, J.S., Brandes, E.R., Schroeder, J.W., Sorenson, R.J., Showalter, H.D., and Simmons,

L.A. (2014). RecO and RecR are necessary for RecA loading in response to DNA damage and replication fork stress. Journal of Bacteriology *196*, 2851–2860.

Leonard, A.C., and Grimwade, J.E. (2005). Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. Molecular Microbiology *55*, 978–985.

Lesterlin, C., Barre, F.-X., and Cornet, F. (2004). Genetic recombination and the cell cycle: What we have learned from chromosome dimers. Molecular Microbiology *54*, 1151–1160.

Lin, M., and Kussell, E. (2019). Inferring bacterial recombination rates from large-scale sequencing datasets. Nature Methods *16*, 199–204.

Linke, R., Limmer, M., Juranek, S.A., Heine, A., and Paeschke, K. (2021). The relevance of g-quadruplexes for dna repair. International Journal of Molecular Sciences *22*, 1–21.

Liu, X., Seet, J.X., Shi, Y., and Bianco, P.R. (2019). Rep and UvrD Antagonize One Another at Stalled Replication Forks and This Is Exacerbated by SSB. ACS Omega *4*, 5180–5196.

Lloyd, R.G. (1983). lexA dependent recombination in uvrD strains of Escherichia coli. MGG Molecular & General Genetics *189*, 157–161.

Lloyd, R.G. (1991). Conjugational recombination in resolvase-deficient ruvC mutants of Escherichia coli K-12 depends on recG. Journal of Bacteriology *173*, 5414–5418.

Lloyd, R.G., and Rudolph, C.J. (2016). 25 years on and no end in sight: a perspective on the role of RecG protein. Current Genetics *62*, 827–840.

Lloyd, R.G., and Sharples, G.J. (1993). Dissociation of synthetic Holliday junctions by E.Coli RecG protein. EMBO Journal *12*, 17–22.

Louarn, J., Patte, J., and Louarn, J.M. (1979). Map position of the replication terminus on the Escherichia coli chromosome. MGG Molecular & General Genetics *172*, 7–11.

Luria, S.E., and Delbrück, M. (1943). MUTATIONS OF BACTERIA FROM VIRUS SENSITIVITY TO VIRUS RESISTANCE. Genetics *28*, 491–511.

Maduike, N.Z., Tehranchi, A.K., Wang, J.D., and Kreuzer, K.N. (2014). Replication of the Escherichia coli chromosome in RNase HI-deficient cells: Multiple initiation regions and fork dynamics. Molecular Microbiology *91*, 39–56.

Maffia, A., Ranise, C., and Sabbioneda, S. (2020). From R-loops to G-quadruplexes: Emerging new threats for the replication fork. International Journal of Molecular Sciences *21*.

Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q., and Lloyd, R.G. (2003). A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. EMBO Journal *22*, 724–734.

Mahdi, A.A., Briggs, G.S., and Lloyd, R.G. (2012). Modulation of DNA damage tolerance in Escherichia coli recG and ruv strains by mutations affecting PriB, the ribosome and RNA polymerase. Molecular Microbiology *86*, 675–691.

Maizels, N., and Gray, L.T. (2013). The G4 Genome. PLoS Genetics *9*.

Mäkelä, J., and Sherratt, D.J. (2020). Organization of the Escherichia coli Chromosome by a MukBEF Axial Core. Molecular Cell *78*, 250–260.e5.

Malig, M., Hartono, S.R., Giafaglione, J.M., Sanz, L.A., and Chedin, F. (2020). Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. Journal of Molecular Biology *432*, 2271–2288.

Malone, E.G., Thompson, M.D., and Byrd, A.K. (2022). Role and Regulation of Pif1 Family Helicases at the Replication Fork. International Journal of Molecular Sciences *23*, 1–20.

Marians, K.J. (2018). Lesion Bypass and the Reactivation of Stalled Replication Forks. Annual Review of Biochemistry *87*, 217–238.

Marinus, M.G., and Løbner-Olesen, A. (2014). DNA Methylation. EcoSal Plus *6*, 12437–12443.

Massey, T.H., Aussel, L., Barre, F.X., and Sherratt, D.J. (2004). Asymmetric activation of Xer site-specific recombination by FtsK. EMBO Reports *5*, 399–404.

Massy, B. de, Patte, J., Louarn, J.M., and Bouché, J.P. (1984). oriX: A new replication origin in E. coli. Cell *36*, 221–227.

Massy, B. de, Béjar, S., Louarn, J., and Bouché, J.P. (1987). Inhibition of replication forks exiting the terminus region of the Escherichia coli chromosome occurs at two loci separated by 5 min. Proceedings of the National Academy of Sciences of the United States of America *84*, 1759–1763.

Masters, M., and Broda, P. (1971). Evidence for the bidirectional replication of the Escherichia coli chromosome. Nature New Biology *232*, 137–140.

Mazoyer, A., Drouilhet, R., Despréaux, S., and Ycart, B. (2017). Flan: An R package for inference on mutation models. R Journal *9*, 334–351.

McGlynn, P., and Lloyd, R.G. (1999). RecG helicase activity at three- and four-strand DNA structures. Nucleic Acids Research *27*, 3049–3056.

McGlynn, P., and Lloyd, R.G. (2000). Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. Cell *101*, 35–45.

McGlynn, P., and Lloyd, R.G. (2001). Rescue of stalled replication forks by RecG: Simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and holliday junction formation. Proceedings of the National Academy of Sciences of the United States of America *98*, 8227–8234.

Melton-Celsa, A.R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. Microbiology Spectrum *2*, 1–21.

Mercier, R., Petit, M.A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., and Espéli, O. (2008). The MatP/matS Site-Specific System Organizes the Terminus Region of the E. coli Chromosome into a Macrodomain. Cell *135*, 475–485.

Merrikh, H. (2017). Spatial and Temporal Control of Evolution through Replication–Transcription Conflicts. Trends in Microbiology *25*, 515–521.

Merrikh, C.N., and Merrikh, H. (2018). Gene inversion potentiates bacterial evolvability and virulence. Nature Communications *9*.

Merrikh, C.N., Weiss, E., and Merrikh, H. (2016). The accelerated evolution of lagging strand genes is independent of sequence context. Genome Biology and Evolution *8*, 3696–3702.

Messer, W. (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiology Reviews *26*, 355–374.

Michel, B., Ehrlich, S.D., and Uzest, M. (1997). DNA double-strand breaks caused by replication arrest. EMBO Journal *16*, 430–438.

Michel, B., Sinha, A.K., and Leach, D.R.F. (2018). Replication Fork Breakage and Restart in Escherichia coli . Microbiology and Molecular Biology Reviews *82*, 1–19.

Midgley-Smith, S.L., Dimude, J.U., Taylor, T., Forrester, N.M., Upton, A.L., Lloyd, R.G., and Rudolph, C.J. (2018). Chromosomal over-replication in Escherichia coli recG cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. Nucleic Acids Research *46*, 7701–7715.

Midgley-Smith, S.L., Dimude, J.U., and Rudolph, C.J. (2019). A role for 3 exonucleases at the final stages of chromosome duplication in Escherichia coli. Nucleic Acids Research.

Miglietta, G., Russo, M., and Capranico, G. (2020). G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. Nucleic Acids Research *48*, 11942–11957.

Moolman, M.C., Tiruvadi Krishnan, S., Kerssemakers, J.W.J., De Leeuw, R., Lorent, V., Sherratt, D.J., and Dekker, N.H. (2016). The progression of replication forks at natural replication barriers in live bacteria. Nucleic Acids Research *44*, 6262–6273.

Morimatsu, K., and Kowalczykowski, S.C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: A universal step of recombinational repair. Molecular Cell *11*, 1337–1347.

Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: Mechanisms and regulation in bacteria. Nature Reviews Microbiology *5*, 343–354.

Mulcair, M.D., Schaeffer, P.M., Oakley, A.J., Cross, H.F., Neylon, C., Hill, T.M., and Dixon, N.E. (2006). A Molecular Mousetrap Determines Polarity of Termination of DNA Replication in E. coli. Cell *125*, 1309–1319.

Mulugu, S., Potnis, A., Shamsuzzaman, Taylor, J., Alexander, K., and Bastia, D. (2001). Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. Proceedings of the National Academy of Sciences of the United States of America *98*, 9569–9574.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006). Bacterial Endosymbiont Carsonella. Science *314*, 267.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability an evolving hallmark of cancer. Nature Reviews Molecular Cell Biology *11*, 220–228.

Newton, K.N., Courcelle, C.T., and Courcelle, J. (2012). UvrD participation in nucleotide excision repair is required for the recovery of DNA synthesis following UV-induced damage in Escherichia coli. Journal of Nucleic Acids *2012*.

Neylon, C., Kralicek, A.V., Hill, T.M., and Dixon, N.E. (2005). Replication Termination in Escherichia coli: Structure and Antihelicase Activity of the Tus-Ter Complex. Microbiology and Molecular Biology Reviews *69*, 501–526.

Oeda, K., Horiuchi, T., and Sekiguchi, M. (1981). Molecular cloning of the uvrD gene of Escherichia coli that controls ultraviolet sensitivity and spontaneous mutation frequency. Molecular and General Genetics MGG *184*, 191–199.

Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P., and Beese, L.S. (2011). Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. Cell *145*, 212–223.

Ordabayev, Y.A., Nguyen, B., Niedziela-Majka, A., and Lohman, T.M. (2018). Regulation of UvrD Helicase Activity by MutL. Journal of Molecular Biology *430*, 4260–4274.

Pandey, M., Elshenawy, M.M., Jergic, S., Takahashi, M., Dixon, N.E., Hamdan, S.M., and Patel, S.S. (2015). Two mechanisms coordinate replication termination by the Escherichia coli Tus-Ter complex. Nucleic Acids Research *43*, 5924–5935.

Parekh, V.J., Węgrzyn, G., Arluison, V., and Sinden, R.R. (2023). Genomic Instability of G-Quadruplex Sequences in Escherichia coli: Roles of DinG, RecG, and RecQ Helicases. Genes *14*, 1720.

Petrova, V., Chen, S.H., Molzberger, E.T., Tomko, E., Chitteni-Pattu, S., Jia, H., Ordabayev, Y., Lohman, T.M., and Cox, M.M. (2015). Active displacement of RecA filaments by UvrD translocase activity. Nucleic Acids Research *43*, 4133–4149.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., and Denamur, E. (1999). The link between phylogeny and virulence in Escherichia coli extraintestinal infection? Infection and Immunity *67*, 546–553.

Pohl, T.J., and Zakian, V.A. (2019). Pif1 family DNA helicases: A helpmate to RNase H? DNA Repair *84*, 102633.

Pomerantz, R.T., and O'Donnell, M. (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. Nature *456*.

Pomerantz, R.T., and O'Donnell, M. (2010). Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. Science *327*, 590–592.

Pupo, G.M., Lan, R., and Reeves, P.R. (2000). Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. Proceedings of the National Academy of Sciences of the United States of America *97*, 10567–10572.

Raghunathan, N., Goswami, S., Leela, J.K., Pandiyan, A., and Gowrishankar, J. (2019). A new role for Escherichia coli Dam DNA methylase in prevention of aberrant chromosomal replication. Nucleic Acids Research *47*, 5698–5711.

Rakowski, S.A., and Filutowicz, M. (2013). Plasmid R6K replication control. Plasmid *69*, 231–242.

Ramirez, P., Crouch, R.J., Cheung, V.G., and Grunseich, C. (2022). R-Loop Analysis by Dot-Blot. 1–16.

Rasko, D.A., Rosovitz, M.J., Myers, G.S.A., Mongodin, E.F., Fricke, W.F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N.R., Chaudhuri, R., et al. (2008). The pangenome structure of Escherichia coli: Comparative genomic analysis of E. coli commensal and pathogenic isolates. Journal of Bacteriology *190*, 6881–6893.

Rehrauer, W.M., Lavery, P.E., Palmer, E.L., Singh, R.N., and Kowalczykowskit, S.C. (1996). Interaction of Escherichia coli RecA protein with LexA repressor. I. LexA repressor cleavage is competitive with binding of a secondary DNA molecule. Journal of Biological Chemistry *271*, 23865–23873.

Reyes-Lamothe, R., Wang, X., and Sherratt, D. (2008). Escherichia coli and its chromosome. Trends in Microbiology *16*, 238–245.

Riber, L., and Løbner-Olesen, A. (2005). Coordinated replication and sequestration of oriC and dnaA are required for maintaining controlled once-per-cell-cycle initiation in Escherichia coli.

Journal of Bacteriology *187*, 5605–5613.

Roecklein, B., Pelletier, A., and Kuempel, P. (1991). The tus gene of Escherichia coli: autoregulation, analysis of flanking sequences and identification of a complementary system in Salmonella typhimurium. Research in Microbiology *142*, 169–175.

Rudolph, C.J., Dhillon, P., Moore, T., and Lloyd, R.G. (2007). Avoiding and resolving conflicts between DNA replication and transcription. DNA Repair *6*, 981–993.

Rudolph, C.J., Upton, A.L., Harris, L., and Lloyd, R.G. (2009). Pathological replication in cells lacking RecG DNA translocase. Molecular Microbiology.

Rudolph, C.J., Mahdi, A.A., Upton, A.L., and Lloyd, R.G. (2010b). RecG protein and single-strand DNA exonucleases avoid cell lethality associated with PriA helicase activity in Escherichia coli. Genetics.

Rudolph, C.J., Upton, A.L., Briggs, G.S., and Lloyd, R.G. (2010c). Is RecG a general guardian of the bacterial genome?

Rudolph, C.J., Mahdi, A.A., Upton, A.L., and Lloyd, R.G. (2010a). RecG protein and single-strand DNA exonucleases avoid cell lethality associated with PriA helicase activity in Escherichia coli. Genetics.

Rudolph, C.J., Upton, A.L., Stockum, A., Nieduszynski, C.A., and Lloyd, R.G. (2013). Avoiding chromosome pathology when replication forks collide. Nature *500*, 608–611.

Rudolph, C.J., Corocher, T.-A., Grainge, I., and Duggin, I.G. (2019). Termination of DNA Replication in Prokaryotes. In eLS, pp.1–15.

Sakai, A., and Cox, M.M. (2009). RecFOR and RecOR as distinct RecA loading pathways. Journal of Biological Chemistry *284*, 3264–3272.

Sandler, S.J., Leroux, M., Windgassen, T.A., and Keck, J.L. (2021). Escherichia coli K-12 has

two distinguishable PriA-PriB replication restart pathways. Molecular Microbiology *116*, 1140–1150.

Santos, J.A., and Lamers, M.H. (2020). Novel antibiotics targeting bacterial replicative dna polymerases. Antibiotics *9*, 1–14.

Sanz, L.A., Castillo-Guzman, D., and Chédin, F. (2021). Mapping r-loops and rna:Dna hybrids with s9.6-based immunoprecipitation methods. Journal of Visualized Experiments *2021*.

Sargentini, N.J., and Smith, K.C. (1981). Much of spontaneous mutagenesis in escherichia coli is due to error-prone DNA repair: Implications for spontaneous carcinogenesis. Carcinogenesis *2*, 863–872.

Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M.O., Bartels, D., Bekel, T., Beyer, S., Bode, E., et al. (2007). Complete genome sequence of the myxobacterium Sorangium cellulosum. Nature Biotechnology *25*, 1281–1289.

Schroeder, J.W., Hirst, W.G., Szewczyk, G.A., and Simmons, L.A. (2016). The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. Current Biology *26*, 692–697.

Schroeder, J.W., Sankar, T.S., Wang, J.D., and Simmons, L.A. (2020). The roles of replication-transcription conflict in mutagenesis and evolution of genome organization. PLoS Genetics *16*, 1–11.

Sherratt, D.J., Arciszewska, L.K., Blakely, G., Colloms, S., Grant, K., Leslie, N., McCulloch, R., Lindahl, T.R., and West, S.C. (1995). Site-specific recombination and circular chromosome segregation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences *347*, 37–42.

Sims, G.E., and Kim, S.H. (2011). Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). Proceedings of the National Academy of Sciences of the United States of America *108*, 8329–8334.

Singleton, M.R., Scaife, S., and Wigley, D.B. (2001). Structural analysis of DNA replication fork reversal by RecG. Cell *107*, 79–89.

Sinha, A.K., Durand, A., Desfontaines, J.M., Iurchenko, I., Auger, H., Leach, D.R.F., Barre, F.X., and Michel, B. (2017). Division-induced DNA double strand breaks in the chromosome terminus region of Escherichia coli lacking RecBCD DNA repair enzyme. PLoS Genetics *13*.

Sinha, A.K., Possoz, C., Durand, A., Desfontaines, J.M., Barre, F.X., Leach, D.R.F., and Michel, B. (2018). Broken replication forks trigger heritable DNA breaks in the terminus of a circular chromosome. PLoS Genetics *14*, 1–28.

Sinha, A.K., Possoz, C., and Leach, D.R.F. (2020). The Roles of Bacterial DNA Double-Strand Break Repair Proteins in Chromosomal DNA Replication. FEMS Microbiology Reviews *44*, 351–368.

Skarstad, K., and Katayama, T. (2013). Regulating DNA replication in bacteria. Cold Spring Harbor Perspectives in Biology *5*, 1–17.

Smith, G.R. (2012). How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination : a Molecular Biologist ' s View RECOMBINATION. *76*, 217–228.

Smolka, J.A., Sanz, L.A., Hartono, S.R., and Chédin, F. (2021). Recognition of rna by the s9.6 antibody creates pervasive artifacts when imaging rna:Dna hybrids. Journal of Cell Biology *220*.

Solar, G. del, Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998). Replication and Control of Circular Bacterial Plasmids. Microbiology and Molecular Biology Reviews *62*, 434–464.

Stanage, T.H., Page, A.N., and Cox, M.M. (2017). DNA flap creation by the RarA/MgsA protein of Escherichia coli. Nucleic Acids Research *45*, 2724–2735.

Steinacher, R., Osman, F., Dalgaard, J.Z., Lorenz, A., and Whitby, M.C. (2012). The DNA helicase

Pfh1 promotes fork merging at replication termination sites to ensure genome stability. Genes and Development *26*, 594–602.

Swings, T., van Den Bergh, B., Wuyts, S., Oeyen, E., Voordeckers, K., Verstrepen, K.J., Fauvart, M., Verstraeten, N., and Michiels, J. (2017). Adaptive tuning of mutation rates allows fast response to lethal stress in escherichia coli. eLife *6*, 1–24.

Syeda, A.H., Dimude, J.U., Skovgaard, O., and Rudolph, C.J. (2020). Too Much of a Good Thing: How Ectopic DNA Replication Affects Bacterial Replication Dynamics (Frontiers Media S.A.).

Thomason, L.C., Costantino, N., and Court, D.L. (2007). E. coli Genome Manipulation by P1 Transduction. Current Protocols in Molecular Biology1.17.1–1.17.8.

Todd, P.A., and Glickman, B.W. (1979). UV protection and mutagenesis in uvrD, uvrE and recL strains of Escherichia coli carrying the pKM101 plasmid. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *62*, 451–457.

Toft, C.J., Moreau, M.J.J., Perutka, J., Mandapati, S., Enyeart, P., Sorenson, A.E., Ellington, A.D., and Schaeffer, P.M. (2021). Delineation of the ancestral tus-dependent replication fork trap. International Journal of Molecular Sciences *22*.

Toft, C.J., Sorenson, A.E., and Schaeffer, P.M. (2022). Rise of the terminator protein tus : A versatile tool in the biotechnologist's toolbox. Analytica Chimica Acta *1213*, 339946.

Tomasetti, C., Li, L., and Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science *355*, 1330–1334.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., et al. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genetics *5*.

Urrutia-Irazabal, I., Ault, J.R., Sobott, F., Savery, N.J., and Dillingham, M.S. (2021). Analysis of the pcra-rna polymerase complex reveals a helicase interaction motif and a role for pcra/uvrd

helicase in the suppression of r-loops. eLife *10*, 1–29.

Vassilev, A., and DePamphilis, M.L. (2017). Links between DNA replication, stem cells and cancer. Genes *8*.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. EMBO Journal *24*, 180–189.

Venkitaraman, A.R. (2019). How do mutations affecting the breast cancer genes BRCA1 and BRCA2 cause cancer susceptibility? DNA Repair *81*, 1–19.

Verma, S.C., Qian, Z., and Adhya, S.L. (2019). Architecture of the Escherichia coli nucleoid.

Vincent, S.D., Mahdi, A.A., and Lloyd, R.G. (1996). The RecG branch migration protein of Escherichia coli dissociates R-loops. Journal of Molecular Biology *264*, 713–721.

Vlachos-Breton, É., and Drolet, M. (2022). R-Loop Detection in Bacteria. pp.31–37.

Wei Dai, Y.Y. (2014). Genomic Instability and Cancer. Journal of Carcinogenesis & Mutagenesis *05*, 1–7.

Wendel, B.M., Courcelle, C.T., and Courcelle, J. (2014). Completion of DNA replication in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *111*, 16454–16459.

Wendel, B.M., Cole, J.M., Courcelle, C.T., and Courcelle, J. (2017). SbcC-SbcD and ExoI process convergent forks to complete chromosome replication. Proceedings of the National Academy of Sciences of the United States of America *115*, 349–354.

Wendel, B.M., Courcelle, C.T., and Courcelle, J. (2020). UV-induced DNA damage disrupts the coordination between replication initiation, elongation, and completion.

White, M.A., Azeroglu, B., Lopez-Vernaza, M.A., Hasan, A.M.M., and Leach, D.R.F. (2018). RecBCD coordinates repair of two ends at a DNA double-strand break, preventing aberrant chromosome amplification. Nucleic Acids Research *46*, 6670–6682.

Willis, N.A., and Scully, R. (2016). Spatial separation of replisome arrest sites influences homologous recombination quality at a Tus/Ter-mediated replication fork barrier. Cell Cycle *15*, 1812–1820.

Willis, N.A., Chandramouly, G., Huang, B., Kwok, A., Follonier, C., Deng, C., and Scully, R. (2014). BRCA1 controls homologous recombination at Tus/Ter-stalled mammalian replication forks. Nature *510*, 556–559.

Willis, N.A., Frock, R.L., Menghi, F., Duffey, E.E., Panday, A., Camacho, V., Hasty, E.P., Liu, E.T., Alt, F.W., and Scully, R. (2017). Mechanism of tandem duplication formation in BRCA1-mutant cells. Nature *551*, 590–595.

Windgassen, T.A., Wessel, S.R., Bhattacharyya, B., and Keck, J.L. (2018). Mechanisms of bacterial DNA replication restart. Nucleic Acids Research *46*, 504–519.

Winterstein, C., and Ludwig, B. (1998). Genes coding for respiratory complexes map on all three chromosomes of the Paracoccus denitrificans genome. Archives of Microbiology *169*, 275–281.

Wollman J., A., Syeda, A.H., A. L. Howard, J., Payne-Dwyer, A., Leech, A., Warecka, D., Guy, C., McGlynn, P., Hawkins, M., and Leake, M.C. (2023). Tetrameric UvrD helicase is located at the E. coli replisome due to frequent replication blocks. Journal of Molecular Biology168369.

Yeom, J., Lee, Y., and Park, W. (2012). ATP-dependent RecG helicase is required for the transcriptional regulator OxyR function in Pseudomonas species. Journal of Biological Chemistry *287*, 24492–24504.

Yokota, H. (2020). DNA-Unwinding Dynamics of Escherichia coli UvrD Lacking the C-Terminal 40 Amino Acids. Biophysical Journal *118*, 1634–1648.

Zenkin, N., Kulbachinskiy, A., Bass, I., and Nikiforov, V. (2005). Different rifampin sensitivities of Escherichia coli and Mycobacterium tuberculosis RNA polymerases are not explained by the difference in the β-subunit rifampin regions I and II. Antimicrobial Agents and Chemotherapy *49*, 1587–1590.

Zhang, C., Joseph, A.M., Casini, L., Collier, J., Badrinarayanan, A., and Manley, S. (2023). Chromosome organization shapes replisome dynamics in Caulobacter crescentus. bioRxiv2023.07.22.550130.

Zheng, Q. (2017). Sample size determination for the fluctuation experiment. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis *795*, 10–14.